

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:57:20 ; Search time 2487.88 Seconds  
(without alignments)  
7673.783 Million cell updates/sec

Title: US-09-581-500B-12

Perfect score: 656

Sequence: 1 gccacacaaacaaatgaaat.....cccacgcgagaccgcgga 656

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pla.\*
- 35: em.htg.rod.\*
- 36: em.htg.man.\*
- 37: em.htg.vit.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	652.6	99.5	656	6	AX021006	AX021006 Sequence
c 2	647.8	98.8	191395	9	AC009802	AC009802 Homo sapi
c 3	646.2	98.5	198291	2	AC067875	AC067875 Homo sapi
c 4	646.2	98.5	201734	2	AC090224	AC090224 Homo sapi
c 5	643	98.0	191793	2	AP001897	AP001897 Homo sapi
c 6	441.6	67.3	13202	6	AX346386	AX346386 Sequence
c 7	376	57.3	186351	2	AP001503	AP001503 Homo sapi
c 8	367.8	56.1	13202	6	AX346387	AX346387 Sequence
c 9	73.4	11.2	1425	6	AR005195	AR005195 Sequence
c 10	73.4	11.2	1425	6	IR1228	IR1228 Sequence 1
c 11	73.4	11.2	1425	6	IR2208	IR2208 Sequence 1
c 12	73	11.1	1325	9	HUMCAP2A	I40377 Homo sapien
c 13	72	11.0	7218	6	I66494	I66494 Sequence 14
c 14	59.6	9.1	113684	2	AC103021	AC103021 Rattus no
c 15	57.4	8.8	65711	2	AC117583	AC117583 Mus muscu
c 16	56.8	8.7	750	11	PM12B11B	AL684309 Penicilli
c 17	55.4	8.4	130051	2	AC127840	AC127840 Rattus no
c 18	55.2	8.4	161310	2	AC120603	AC120603 Rattus no
c 19	54.8	8.4	156782	2	AC114101	AC114101 Rattus no
c 20	54	8.2	168542	2	AC058792	AC058792 Homo sapi
c 21	54	8.2	189218	9	AL355297	AL355297 Human DNA
c 22	53.8	8.2	83862	2	AC120671	AC120671 Rattus no
c 23	53.6	8.2	6170	9	HSCGSE7	AL672255 Mus muscu
c 24	53.6	8.2	169232	2	AC129646	AC129646 Rattus no
c 25	53.6	8.2	245249	2	AC103134	AC103134 Rattus no
c 26	53.4	8.1	103874	2	AC106598	AC106598 Rattus no
c 27	53.4	8.1	194260	2	AP005508	AP005508 Oryza sat
c 28	53	8.1	7450	9	AB035726	AB035726 Homo sapi
c 29	53	8.1	145489	2	AL672255	AL672255 Mus muscu
c 30	53	8.1	220201	10	AL669898	AL669898 Mouse DNA
c 31	52.8	8.0	137721	2	AC113928	AC113928 Rattus no
c 32	52.6	8.0	119825	2	AC128152	AC128152 Rattus no
c 33	52.6	8.0	228584	2	AL669838	AL669838 Mus muscu
c 34	52.4	8.0	150653	2	AC109865	AC109865 Rattus no
c 35	52	7.9	922	9	HS092984	U92984 Homo sapien
c 36	52	7.9	1319	9	BC034528	BC034528 Homo sapi
c 37	52	7.9	193446	2	AC094597	AC094597 Rattus no
c 38	51.8	7.9	229380	2	AC079636	AC079636 Mus muscu
c 39	51.6	7.9	125020	9	AF429315	AF429315 Homo sapi
c 40	51.6	7.9	139838	2	AC119485	AC119485 Rattus no
c 41	51.4	7.8	108019	2	AC105476	AC105476 Rattus no
c 42	51.4	7.8	150074	2	AC016007	AC016007 Homo sapi
c 43	51.4	7.8	151729	2	AC124866	AC124866 Rattus no
c 44	51	7.8	76118	2	AC094984	AC094984 Rattus no
c 45	50.6	7.7	259	9	AB018491	AB018491 Homo sapi

ALIGNMENTS

RESULT 1  
AX021006  
LOCUS AX021006 656 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 12 from Patent WO9932643.  
ACCESSION AX021006  
VERSION AX021006.1 GI:10044669  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 656)  
AUTHORS Del-Favero,J., Raeymaekers,P. and Van Broeckhoven,C.  
TITLE Mood disorder gene  
JOURNAL Patent: WO 9932643-A 12 01-JUL-1999;

BROECKHOVEN CHRISTINE VAN (BE); DEL FAVERO JURGEN (BE); RAYTEMARKERS  
PETER (BE); VIAAMS INTERUNIV INST BIOTECH (BE)

FEATURES  
source  
Location/Qualifiers  
1. .656  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT  
183 a 165 c 208 g 96 t 4 others

ORIGIN

Query Match 99.5%; Score 652.6; DB 6; Length 656;  
Best Local Similarity 100.0%; Pred. No. 8.1e-148; Indels 0; Gaps 0;  
Matches 656; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCACAAACAAATGAATGAACCTGGATGTATTTTGGCCAGCAATAGAAAA 60  
Db 1 GCACAAACAAATGAATGAACCTGGATGTATTTTGGCCAGCAATAGAAAA 60

QY 61 TGATTAGTATCCTTATCAGGAGCAATTCAGAGATGTTGGTGGACCTCAACTACAG 120  
Db 61 TGATTAGTATCCTTATCAGGAGCAATTCAGAGATGTTGGTGGACCTCAACTACAG 120

QY 121 TGGAGTCAACGCTGAATCAACGCTGAAAGAGCAATAGCAATGTGACACTTTTAT 180  
Db 121 TGGAGTCAACGCTGAATCAACGCTGAAAGAGCAATAGCAATGTGACACTTTTAT 180

QY 181 AAAAACCACCTCCAGGACCAAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240  
Db 181 AAAAACCACCTCCAGGACCAAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240

QY 241 CAGGCCCAAAATCAACGCTGATTCACAGCAGCAATGCAAGCTTCTGATCATCTG 300  
Db 241 CAGGCCCAAAATCAACGCTGATTCACAGCAGCAATGCAAGCTTCTGATCATCTG 300

QY 301 CCGAGAGCTTGCCCTCAATCAGGCGGACGTAAGCACTATCAAGAGGAGGATGACAA 360  
Db 301 CCGAGAGCTTGCCCTCAATCAGGCGGACGTAAGCACTATCAAGAGGAGGATGACAA 360

QY 361 AGCAGCAGCGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
Db 361 AGCAGCAGCGGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420

QY 421 CAGAGTACCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
Db 421 CAGAGTACCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480

QY 481 CTGGAGTGGCTGAGAGAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
Db 481 CTGGAGTGGCTGAGAGAGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540

QY 541 GCGGGCGAGAACTANGAATGACGCGGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
Db 541 GCGGGCGAGAACTANGAATGACGCGGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600

QY 601 GTGGGGTCCCTCGCAGGCGGGGCTCCCAAGCCCGGCGGAGCCCGCGGAGCCCGCGGA 656  
Db 601 GTGGGGTCCCTCGCAGGCGGGGCTCCCAAGCCCGGCGGAGCCCGCGGAGCCCGCGGA 656

RESULT 2  
AC009802/c  
LOCUS  
DEFINITION Homo sapiens chromosome 18, clone RP11-793J2, complete sequence.  
ACCESSION AC009802  
VERSION AC009802.13 GI:6705901  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 191395)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 18, clone RP11-793J2  
JOURNAL unpublished

2 (bases 1 to 191395)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Nilloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Torriella-Miller,I., Vassiliev,H., Vo.A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

Direct Submission  
Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 191395)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,  
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepl,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,  
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,  
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Triggilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.

Direct Submission  
Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jan 16, 2000 this sequence version replaced gi:6514043.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1250  
Center clone name: 793\_J2  
----- Location/Qualifiers  
1. .191395  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-793J2"  
/clone\_lib="RP11-11 Human Male BAC"  
2. .542  
/rpt\_family="L1M2"  
474. .708  
/rpt\_family="L1"  
708. .1353  
/rpt\_family="L1"  
1351. .1623  
/rpt\_family="L1M4A"

repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region

FEATURES  
source  
Location/Qualifiers  
1. .191395  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-793J2"  
/clone\_lib="RP11-11 Human Male BAC"  
2. .542  
/rpt\_family="L1M2"  
474. .708  
/rpt\_family="L1"  
708. .1353  
/rpt\_family="L1"  
1351. .1623  
/rpt\_family="L1M4A"

repeat_region	1632..1687	/rpt_family="TATATG)n"
unsure		complement(1673..1726)
repeat_region	1688..2207	/note="Single-stranded coverage."
unsure		/rpt_family="L1M4A4"
repeat_region		complement(1724..1770)
repeat_region	2208..2215	/note="Single-stranded terminator coverage."
repeat_region	2594..2648	/rpt_family="TATATG)n"
repeat_region		/rpt_family="L1ME"
repeat_region		complement(2649..2991)
repeat_region	2992..3475	/rpt_family="MER2"
repeat_region		/rpt_family="L1ME"
repeat_region		complement(4330..4437)
repeat_region	4438..4567	/rpt_family="FRAM"
repeat_region		/rpt_family="(TA)n"
repeat_region	4599..4959	/rpt_family="AT_rich"
repeat_region	4982..5151	/rpt_family="(TATG)n"
repeat_region	5144..5311	/rpt_family="(TATAA)n"
repeat_region	5257..5441	/rpt_family="(TA)n"
repeat_region	5444..5525	/rpt_family="AT_rich"
repeat_region	5533..5708	/rpt_family="(TTATA)n"
repeat_region	5742..5907	/rpt_family="(TTATA)n"
repeat_region		complement(5966..6047)
repeat_region	6682..6703	/rpt_family="FRAM/FRAN"
repeat_region	7034..7933	/rpt_family="AT_rich"
repeat_region	7935..8006	/rpt_family="L1PA16"
repeat_region	8009..9111	/rpt_family="AluY"
repeat_region	9112..9165	/rpt_family="L1PA16"
repeat_region	9216..9496	/rpt_family="AT_rich"
repeat_region	9497..9520	/rpt_family="AluX"
repeat_region	9961..10264	/rpt_family="(CAA)n"
unsure		complement(10837..10840)
repeat_region		/note="Single-stranded terminator coverage."
repeat_region	11856..11987	complement(11156..11515)
repeat_region	12592..12643	/rpt_family="HEIC"
repeat_region	13683..13994	/rpt_family="L12"
repeat_region		/rpt_family="AluX"
repeat_region		complement(13996..14160)
repeat_region	14788..15255	/rpt_family="LINEC"
repeat_region		/rpt_family="MER67D"
repeat_region		complement(15875..15996)
repeat_region	16101..16135	/rpt_family="MERSB"
repeat_region	16136..16304	/rpt_family="AT_rich"
repeat_region		/rpt_family="L1M4"
repeat_region		complement(16572..16647)

[illegible]



```

RESULT 4
AC090224/c
LOCUS
DEFINITION
  Homo sapiens chromosome 18 clone RP11-693A18 map 18, WORKING DRAFT
  SEQUENCE, 8 unordered pieces.
AC090224
VERSION
  AC090224.4 GI:14190716
KEYWORDS
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
  Homo sapiens
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 201734)
  Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 18, clone RP11-693A18
  Unpublished
  2 (bases 1 to 201734)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
  Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
  Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
  Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
  Dodge,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
  Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
  Haggis,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
  Jones,C., Karatas,A., LaRoque,K., Lamazares,R., Landers,T.,
  Lebecky,J., Levine,R., Liu,G., MacLean,C., MacDonald,P.,
  Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
  McPheters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
  Murphy,T., Nguyen,C., Norbu,C., Norman,C.H.,
  O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
  Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
  Rieback,M., Riley,R., Rise,C., Roqop,P., Roman,J., Rosetti,M.,
  Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
  Sounez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
  Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
  Travers,M., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
  Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
  Zembek,L., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On May 23, 2001 this sequence version replaced gi:13621272.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu
  ----- Project Information
  Center project name: L12265
  Center clone name: 693_A_18
  ----- Summary Statistics
  Sequencing vector: Plasmid; n/a; 100% of reads
  Chemistry: Dye-terminator Big Dye; 100% of reads
  Assembly program: Phrap; version 0.960731
  Consensus quality: 198542 bases at least Q40
  Consensus quality: 200270 bases at least Q30
  Consensus quality: 200829 bases at least Q20
  Insert size: 194000; agarose-fp
  Insert size: 201034; sum-of-contigs
  Quality coverage: 11.3 in Q20 bases; agarose-fp
  Quality coverage: 10.9 in Q20 bases; sum-of-contigs
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 8 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will

```

```

* be preserved.
* 1 45183: contig of 45183 bp in length
* 45184 45283: gap of 100 bp
* 45284 46768: contig of 1485 bp in length
* 46769 46868: gap of 100 bp
* 46869 47961: contig of 1093 bp in length
* 47962 48061: gap of 100 bp
* 48062 49643: contig of 1582 bp in length
* 49644 49743: gap of 100 bp
* 49744 51393: contig of 1650 bp in length
* 51394 51493: gap of 100 bp
* 51494 74531: contig of 23038 bp in length
* 74532 74631: gap of 100 bp
* 74632 127710: contig of 53079 bp in length
* 127711 127810: gap of 100 bp
* 127811 201734: contig of 73924 bp in length.

FEATURES
    Location/Qualifiers
        1..201734
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="18"
            /map="18"
            /clone="RP11-693A18"
            /clone_lib="RPC1-11 Human Male BAC"
        misc_feature
            1..45183
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left
                45284..46768
                    /note="assembly_fragment"
                    46869..47961
                        /note="assembly_fragment"
                        48062..49643
                            /note="assembly_fragment"
                            49744..51393
                                /note="assembly_fragment"
                                51494..74531
                                    /note="assembly_fragment"
                                    74632..127710
                                        /note="assembly_fragment"
                                        127811..201734
                                            /note="assembly_fragment"
                                            clone_end:17
                                            vector_side:right
                                            61611 a 39444 c 39572 g 60399 t 708 others
                                            ORIGIN
Query Match          98.5%; Score 646.2; DB 2; Length 201734;
Best Local Similarity 98.8%; Pred. No. 3.8e-146;
Matches 648; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 GCCAACAAACAAATGAATAGACCTGGGATGTAATTTTGGCCAAAGGCAATAGAAAA 60
    |
    |
    |
Db 148069 GCCAACAAACAAATGAATAGACCTGGGATGTAATTTTGGCCAAAGGCAATAGAAAA 148010
    |
    |
    |
QY 61 TGATTAGTATCCTTATCAGGAGCAATTTAGAGCAATTTGGTGGAGCTGTAACACAG 120
    |
    |
    |
Db 148009 TGATTAGTATCCTTATCAGGAGCAATTTAGAGCAATTTGGTGGAGCTGTAACACAG 147950
    |
    |
    |
QY 121 TGGAGTCAACAGCTGAATCAACGGTGAAGGAGCAATAGGCAATAGTGTACACTTTTAT 180
    |
    |
    |
Db 147949 TGGAGTCAACAGCTGAATCAACGGTGAAGGAGCAATAGGCAATAGTGTACACTTTTAT 147890
    |
    |
    |
QY 181 AAAAACCCCTCCCAAGGAGCAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 240
    |
    |
    |
Db 147889 AAAARCCACCTCCCAAGGAGCAGGCACTGGCCCTCTCTCCGGTGCCACAGACATCCACA 147830
    |
    |
    |
QY 241 CAGGCCCAAGATCAGGATTCAGCAAGCCAGAGCAATCGAACGGTCTCTGATGATCTG 300
    |
    |
    |
Db 147829 CAGGCCCAAGATCAGGATTCAGCAAGCCAGAGCAATCGAACGGTCTCTGATGATCTG 147770
    |
    |
    |
QY 301 CCGGAGCCTTCCTCAATCAATCAGCGGAGCTGAGCATCTACAGGAGGAGTACTCAA 360
    |
    |
    |

```

preserved	1	10723	contig of	10723	bp in length
	10824	19450	contig of	8627	bp in length
	19551	28739	contig of	9189	bp in length
	28840	37796	contig of	8957	bp in length
	37897	45643	contig of	7747	bp in length
	45744	53200	contig of	7457	bp in length
	53301	59466	contig of	6166	bp in length
	59567	65023	contig of	5457	bp in length
	65124	71725	contig of	6602	bp in length
	71826	77046	contig of	5221	bp in length
	77147	82059	contig of	4913	bp in length
	82160	86515	contig of	4356	bp in length
	86616	90836	contig of	4221	bp in length
	90937	95257	contig of	4321	bp in length
	95358	99276	contig of	3919	bp in length
	99377	103751	contig of	4375	bp in length
	103852	108407	contig of	4556	bp in length
	108508	112468	contig of	3961	bp in length
	112569	116411	contig of	3843	bp in length
	116512	120690	contig of	4179	bp in length
	120791	125789	contig of	4999	bp in length
	125890	129429	contig of	3540	bp in length
	129530	132875	contig of	3346	bp in length
	132976	136718	contig of	3743	bp in length
	136819	140204	contig of	3386	bp in length
	140305	143212	contig of	2908	bp in length
	143313	146311	contig of	2999	bp in length
	146412	148694	contig of	2283	bp in length
	148795	151551	contig of	2757	bp in length
	151652	153471	contig of	1820	bp in length
	153572	155905	contig of	2334	bp in length
	156006	157930	contig of	1925	bp in length
	158031	160335	contig of	2305	bp in length
	160436	161957	contig of	2052	bp in length
	162058	164115	contig of	2058	bp in length
	164216	166374	contig of	2159	bp in length
	166475	168634	contig of	2160	bp in length
	169735	170796	contig of	2062	bp in length
	170897	172731	contig of	1835	bp in length
	172832	174464	contig of	1633	bp in length
	174565	175680	contig of	1116	bp in length
	175781	176918	contig of	1138	bp in length
	177019	178822	contig of	1804	bp in length
	178923	180077	contig of	1155	bp in length
	180178	181248	contig of	1071	bp in length
	181349	182439	contig of	1151	bp in length
	182600	182806	contig of	207	bp in length
	182907	183932	contig of	1086	bp in length
	184093	185198	contig of	1106	bp in length
	185299	186320	contig of	1022	bp in length
	186421	187601	contig of	1181	bp in length
	187702	189039	contig of	1398	bp in length
	189200	190576	contig of	1377	bp in length
	190677	191793	contig of	1117	bp in length

\* 10723: contig of 10723 bp in length  
\* 10724 10823: gap of 100 bp  
\* 10824 19450: contig of 8627 bp in length  
\* 19451 19550: gap of 100 bp  
\* 19551 28739: contig of 9189 bp in length  
\* 28740 28839: gap of 100 bp  
\* 28840 37796: contig of 8957 bp in length

```

----- genome center
Center: RIKEN Genomic Sciences Center(gsc)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-693A18
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167461 bases at least Q40
Consensus quality: 174712 bases at least Q30
Consensus quality: 180769 bases at least Q20
Insert size: 186493; sum-of-contigs
Quality coverage: 4.60x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps

	* 174565	175680:	contig of 1116 bp in length	
	* 175681	175780:	gap of 100 bp	
	* 175781	176918:	contig of 1138 bp in length	
	* 176919	177018:	gap of 100 bp	
	* 177019	178922:	contig of 1804 bp in length	
	* 178923	178922:	gap of 100 bp	
Query Match		98.0%;	Score 643; DB 2; Length 191793;	
Best Local Similarity		98.5%;	Pred. No. 2.3e-145;	
Matches	646;	Conservative	1; Mismatches	9; Indels
			0; Gaps	
QY	1	GCCACAAACAATAATGAATATAGACTCGGATGTATTTTGGCCAACGCATTAGAAAA	60	
Dd	28051	GCCAACAACAATAATGAATATAGACTCGGATGTATTTTGGCCAACGCATTAGAAAA	281110	
QY	61	TGATTAGTATTCCTTATCAGGACCAATTTTCAGAGAATGTTTGSGTGGAGCTCTAACTACAG	120	
Dd	28111	TGATTAGTATTCCTTATCAGGACCAATTTTCAGAGAATGTTTGSGTGGAGCTCTAACTACAG	281707	
QY	121	TGSAGTCAAACGTGAATCAACGCTGAAAAGAGCAATGCCAATGTGTACACTTTTTAT	180	
Dd	28171	TGSAGTCAAACGTGAATCAACGCTGAAAAGAGCAATGCCAATGTGTACACTTTTTAT	28230	
QY	181	AAAAACCACCTCCAGGACCAAGGCATGCGCCCTCTCTCGGTGGCCACAGACATCCACA	240	
Dd	28231	AAAAACCACCTCCAGGACCAAGGCATGCGCCCTCTCTCGGTGGCCACAGACATCCACA	28290	
QY	241	CAGGCCCAAGAANTCAGGGATGTCACAGCCAGACGACATCGAACGGTCTGAGTCATCTG	300	
Dd	29291	CAGGCCCAAGAANTCAGGGATGTCACAGCCAGACGACATCGAACGGTCTGAGTCATCTG	28350	
QY	301	CCGGAGGCTTGGCCCTCAATCAAGCGGACGCTGAAGCATCTCAAGAGGAGGAATAGTCAA	360	
Dd	28351	CCGGAGGCTTGGCCCTCAATCAAGCGGACGCTGAAGCATCTCAAGAGGAGGAATAGTCAA	28410	
QY	361	AGCAGCAGCGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGGCTGGGGCCCTCTG	420	
Dd	28411	AGCAGCAGCGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGGAGGTGGGGCCCTCTG	28470	
QY	421	CCAGGTACCGGGCGGCGCAGCAGGAGTGCACAGTTCGCCAGTTCGCCGCGGACCACTCTTCC	480	
Dd	28471	CCAGGTACCGGGCGGCGCAGCAGGAGTGCACAGTTCGCCAGTTCGCCGCGGAGCCACTCTTCC	28533	
QY	481	CTGGAGTCCGTGAGAGAGGGAGGAGGAGGCCACAGCAGAGCAATCAGCAGCGAGGCAAA	540	
Dd	28531	CTGGAGTCCGTGAGAGAGGGAGGAGGAGGAGGCCACAGCAGAGCAATCAGCAGCGAGGCAAA	28590	
QY	541	GGCGGCGAGGAACCTANGAGAATGACGCGGAGGAGCGCGCGGGAAGAAANTCTCGGGGCT	600	
Dd	28591	GGCGGCGAGGAACCTANGAGAATGACGCGGAGGAGCGCGCGGGAAGAAAGAGAGTCTCGGGTCT	28650	
QY	601	GTGGGGGTTCNCCCTGGCACCCAGCCGGGGTTCCTCAAGCCCCACCGCAGACCCCOCGGA	656	
Dd	28651	GTGGGGGTTCNCCCTGGCACCCAGCCGGGGTTCCTCAAGCCCCACCGCAGACCCCOCGGA	28706	

RESULT 6	AX346386	13202 bp	DNA	linear	PAT 01-FEB-2002
LOCUS	AX346386	Sequence	1457 from Patent WO0200928.		
DEFINITION	AX346386	Accession			
ACCESSION	AX346386	GI:18494272			
VERSION	AX346386.1				
KEYWORDS	.	synthetic construct.			
WORDS	SOURCE	synthetic construct			
ORGANISM	SOURCE	artificial sequences.			
REFERENCE	1				
AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.				
TITLE	Diagnosis of diseases associated with the immune system				
JOURNAL	Patent: WO 0200928-A 1457 03-JAN-2002;				
FEATURES	Epigenomics AG (DE)				
SOURCE	Location/Qualifiers				
	1. .13202				

```
/organism="synthetic construct"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3470 a 187 c 2897 g 6648 t
ORIGIN
Query Match      67.3%; Score 441.6; DB 6; Length 13202;
Best Local Similarity 79.5%; Pred. No. 1.5e-96;
Matches 519; Conservative 1; Mismatches 133; Indels 0; Gaps 0;
OY 4 AACAAACAAATGAATGAACCTGGGATGATATTTTGGCCAAAGCAATAGAAAATGA 63
Db 4670 AATAAATAAATGAATGAATGGGATGATATTTTGGTAAAGTAAATGAATAATGA 4729
OY 64 TTAGTATCCTTATCAGGAGCAATTCAGAGCAATGTTTGGGTGGACGTCACATCAGTGG 123
Db 4730 TTAGTATTTTATAGGAGTAATTTAGAGCAATGTTTGGGTGGACGTCACATCAGTGG 4789
OY 124 AGTCAACCTGATCAACGCTGAAAGAGCAATAGCAATGTTGATACATTTTATATAA 183
Db 4790 AGTTAAACCTGATTAACGCTGAAAGAGCAATAGTATGATATTTTATATAA 4849
OY 184 AACACCTCCAAAGACACGACCTGGCCCTCTCCCGTCCGACACAGACATCCACAG 243
Db 4850 AATTATTTTAAAGATTAGTATGCTGTTTTTTTTCGGTCTTTATAGATATTTATAG 4909
OY 244 GCCCAAGCAATCAGGATTGACACAGCAGCAATCAACGCTTCGTAGTCATCTGCG 303
Db 4910 GTTAAAGATTAGGGAATGTAAGTTAGTAGTATCAACGCTTTGAGTTATTTGTCG 4969
OY 304 GAAGCCTTCCTCAATCAAGCGGAGTGGAAGCATCTACAAAGAGGAATAGTCAAAG 363
Db 4970 GAAGTTTCTGTTTAAATTAAGCGGACGCTCAAGTATTTTAAAGGAGGAATAGTAAAGT 5029
OY 364 AGCAGCGCGCGCGCGCGCGGACGACGACGACGACGACGAGTGGGCGCTCGCA 423
Db 5030 AGTAGCGCGCGCGCGCGCGGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGT 5089
OY 424 GGTACCGCGCGCGCGCGGACGAGGTGCCAGGTTCGCCGGAGGACCTCTCCCTG 483
Db 5090 GGTATCGCGCGGTAGTACGAGGTGTTAGTTTTCGCGGAGTTATTTTTTTTTTG 5149
OY 484 GAGTGGCTGAGAGGGAAGGAGGAGGACGACGAGGAGGATCAGAGCAGGACCAAGGC 543
Db 5150 GAGTGGCTGAGAGGGAAGGAGGAGGAGGAGGATAGAGTAGGAAATAGAGCAGGATAAGGC 5209
OY 544 GGGCAGCACTANGAAGATGATCSCGGGAGGCGCGGCGGAGGAAATCTCGGGGCTGTG 603
Db 5210 GGGTAGGAATTAGGAGATGACGCGGAGGCGGTGCGGGAAGAGAGTCCGGGGTGTG 5269
OY 604 GGGGTCCCTCGCACACCGCGGGGTCCTCCAGCCCGACCGGAGACCCCGCA 656
Db 5270 GGGGTCTGTTTGGTATTAGTCCGGGTTTTAAGTTTATTCGGGAGATTTCGCGCA 5322

RESULT 7
AP001503
LOCUS      AP001503      186351 bp      DNA      linear      HTG 30-MAY-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-879N20 map 18q22, WORKING
DRAFT SEQUENCE, 23 unordered pieces.
ACCESSION      AP001503
VERSION        AP001503.2 GI:8117361
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 186351)
AUTHORS        Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,
              Fujiyama.A., Iada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.
TITLE          Homo sapiens 186,351 genomic DNA of 18q22
JOURNAL        Published Only in DataBase (2000)
REFERENCE      2 (bases 1 to 186351)
```

AUTHORS  
TITLE  
JOURNAL

Hattori.M., Ishii.K., Toyoda.A., Taylor.T.D., Hong-Seog.P.,  
Fujiyama.A., Iada.T., Totoki.Y., Watanabe.H. and Sakaki.Y.  
Direct Submission  
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,  
Japan (E-mail:hattori@gscc.riken.go.jp,  
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,  
Fax:81-42-778-9924)

## COMMENT

On May 30, 2000 this sequence version replaced gi:7288199.  
----- Genome Center  
Center: RIKEN Genomic Sciences Center(GSC)  
Center code: RIKEN  
Web site: http://hgp.gsc.riken.go.jp/  
Contact: hattori@gscc.riken.go.jp  
----- Project Information  
Center project name: HumDraft18  
Center clone name: RP11-879N20  
----- Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 162162 bases at least Q40  
Consensus quality: 175389 bases at least Q30  
Consensus quality: 181326 bases at least Q20  
Insert size: 184151; sum-of-contigs  
Quality coverage: 4.54x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of  
23 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

1	4519	contig of	42519	bp in length
	5952	contig of	17233	bp in length
	74496	contig of	14544	bp in length
	74597	contig of	12844	bp in length
	87541	contig of	12037	bp in length
	99678	contig of	10982	bp in length
	110659	contig of	4654	bp in length
	115413	contig of	9477	bp in length
	11760	contig of	10654	bp in length
	115514	contig of	10128	bp in length
	125091	contig of	7098	bp in length
	135744	contig of	3128	bp in length
	145972	contig of	5226	bp in length
	153170	contig of	3864	bp in length
	156398	contig of	3910	bp in length
	153271	contig of	3756	bp in length
	161724	contig of	2110	bp in length
	156499	contig of	1439	bp in length
	161825	contig of	1694	bp in length
	165688	contig of	2439	bp in length
	165898	contig of	1908	bp in length
	169698	contig of	1017	bp in length
	169799	contig of	1490	bp in length
	173554	contig of		
	173655	contig of		
	175865	contig of		
	177303	contig of		
	177404	contig of		
	179097	contig of		
	181636	contig of		
	181737	contig of		
	183644	contig of		
	183745	contig of		
	184761	contig of		
	184862	contig of		
	186351	contig of		

Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	42519:	contig of	42519	bp in length
*	42520	42619:	gap of	100 bp
*	42620	59852:	contig of	17233 bp in length
*	59853	59952:	gap of	100 bp
*	59953	74496:	contig of	14544 bp in length
*	74497	74596:	gap of	100 bp



```

* 74597 87440: contig of 12844 bp in length
* 87441 87540: gap of 100 bp
* 87541 99577: contig of 12037 bp in length
* 99578 110659: contig of 10982 bp in length
* 110660 110759: gap of 100 bp
* 110760 115413: contig of 4654 bp in length
* 115414 115513: gap of 100 bp
* 115514 124990: contig of 9477 bp in length
* 124991 125090: gap of 100 bp
* 125091 135744: contig of 10654 bp in length
* 135745 135844: gap of 100 bp
* 135845 145972: contig of 10128 bp in length
* 145973 146072: gap of 100 bp
* 146073 153170: contig of 7098 bp in length
* 153171 153270: gap of 100 bp
* 153271 156399: contig of 3128 bp in length
* 156400 156499: gap of 100 bp
* 156500 161724: contig of 5226 bp in length
* 161725 161824: gap of 100 bp
* 161825 165688: contig of 3864 bp in length
* 165689 165788: gap of 100 bp
* 165789 169698: contig of 3910 bp in length
* 169699 169798: gap of 100 bp
* 169799 173554: contig of 3756 bp in length
* 173555 173654: gap of 100 bp
* 173655 175764: contig of 2110 bp in length
* 175765 175864: gap of 100 bp
* 175865 177303: contig of 1439 bp in length
* 177304 177403: gap of 100 bp
* 177404 179097: contig of 1694 bp in length
* 179098 179197: gap of 100 bp
* 179198 181636: contig of 2439 bp in length
* 181637 181736: gap of 100 bp
* 181737 183644: contig of 1908 bp in length
* 183645 183744: gap of 100 bp
* 183745 184761: contig of 1017 bp in length
* 184762 184861: gap of 100 bp
* 184862 186351: contig of 1490 bp in length.

```

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q22"
/clone="RP11-879N20"

```

```

misc_feature /note="assembly_fragment"
misc_feature 42620..59852
misc_feature /note="assembly_fragment"
misc_feature 59953..74496
misc_feature /note="assembly_fragment"
misc_feature 74597..87440
misc_feature /note="assembly_fragment"
misc_feature 87541..99577
misc_feature /note="assembly_fragment"
misc_feature 99678..110659
misc_feature /note="assembly_fragment"
misc_feature 110760..115413
misc_feature /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 115514..124990
misc_feature /note="assembly_fragment"
misc_feature 125091..135744
misc_feature /note="assembly_fragment"
misc_feature 135845..145972
misc_feature /note="assembly_fragment"
misc_feature 146073..153170
misc_feature /note="assembly_fragment"
misc_feature 153271..156398
misc_feature /note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature 156499..161724
misc_feature /note="assembly_fragment"
misc_feature 161825..165688

```

```

misc_feature /note="assembly_fragment"
misc_feature 165789..169698
misc_feature /note="assembly_fragment"
misc_feature 169799..173554
misc_feature /note="assembly_fragment"
misc_feature 173655..175764
misc_feature /note="assembly_fragment"
misc_feature 175865..177303
misc_feature /note="assembly_fragment"
misc_feature 177404..179097
misc_feature /note="assembly_fragment"
misc_feature 179198..181636
misc_feature /note="assembly_fragment"
misc_feature 181737..183644
misc_feature /note="assembly_fragment"
misc_feature 183745..184761
misc_feature /note="assembly_fragment"
misc_feature 184862..186351
misc_feature /note="assembly_fragment"

BASE COUNT 56377 a 36459 c 35952 g 55363 t 2200 others
ORIGIN

Query Match 57.3%; Score 376; DB 2; Length 186351;
Best Local Similarity 96.7%; Pred. No. 1.5e-80;
Matches 406; Conservative 0; Mismatches 10; Indels 4; Gaps 2;

Qy 1 GCCACAAACAAATGAATGAACCTGGATGATATTTTGGCCAAAGCAATTAGAAA 60
Db 87021 GCCACAAACAAATGAATGAACCTGGATGATATTTTGGCCAAAGCAATTAGAAA 60

Qy 61 TGATAGTATCTTATCAGGAGCAATTCAGAGATGTTGGTGGACGCTTAACAG 120
Db 87081 TGATAGTATCTTATCAGGAGCAATTCAGAGATGTTGGTGGACGCTTAACAG 120

Qy 121 TGGAGTCAACAGTGAATCAACGCTGGAAGAGCAATAGCCATGCTACACTTTTAT 180
Db 87141 TGGAGTCAACAGTGAATCAACGCTGGAAGAGCAATAGCCATGCTACACTTTTAT 180

Qy 181 AAAACACCCCTCCCAAGGACGAGCAGCTGGCCCTCTCCGGTGCCCAAGACATCCACA 240
Db 87201 AAAACACCCCTCCCAAGGACGAGCAGCTGGCCCTCTCCGGTGCCCAAGACATCCACA 240

Qy 241 CAGGCCCAAGATCAGGATTCACAAGCCAGAGCAATGACCGTTCTGATCATCTG 300
Db 87261 CAGGCCCAAGATCAGGATTCACAAGCCAGAGCAATGACCGTTCTGATCATCTG 300

Qy 301 CCGGAAGCCTTGCCCTCAATCAAGCGG---ACGTGAAGCATCTACAAAGAGGAATAGT 357
Db 87321 CCGGAAGCCTTGCCCTCAATCAAGCGGCGGACGCTTGAATGATCTACAAAGAGGAATAGT 357

Qy 358 CAAAG-CAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 416
Db 87381 CAAAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 416

RESULT 8
AX346387/c AX346387 13202 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 1458 from Patent WO0200928.
DEFINITION AX346387
ACCESSION AX346387
VERSION AX346387.1 GI:18494273
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1
AUTHORS Olek A., Piepenbrock C. and Berlin K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1458 03-JAN-2002;
FEATURES Epigenomics AG (DE)
Location/Qualifiers
1..13202
/organism="synthetic construct"

```

```
/db xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3984 a 187 c 2851 g 6180 t
ORIGIN
Query Match 56.1%; Score 367.8; DB 6; Length 13202;
Best Local Similarity 72.3%; Pred. No. 1.2e-78;
Matches 474; Conservative 1; Mismatches 181; Indels 0; Gaps 0;
QY 1 GCCACAAACAAATGAATGAACCTGGGATGATTTTGGCCGACGCAATAGAAAA 60
Db 8536 GCCAACAAACAAATGAATGAACCTGGGATGATTTTGGCCGACGCAATAGAAAA 8477
QY 61 TGATTAGTATCCTTATCAGCAGCAATTCAGAGATGTTGGTGACGCTCTAAGTACAG 120
Db 8476 TTAATTAATCCTTATCAGCAGCAATTCAGAGATGTTGGTGACGCTCTAAGTACAG 8417
QY 121 TGGAGTCAAGAGTGAATCAACGGTGAAGAAAGAGCAATAGCAATGTGACACTTTTAT 180
Db 8416 TAAATCAAGAGTGAATCAACGGTGAAGAAAGAGCAATAGCAATGTGACACTTTTAT 8357
QY 181 AAAAGCACCCTCCAGAGCAGCAGGACTGGCCCTCTCTCCGGTCCCGACAGACATCCACA 240
Db 8356 AAAAGCACCCTCCAGAGCAGCAGGACTGGCCCTCTCTCCGGTCCCGACAGACATCCACA 8297
QY 241 CAGGCCCAAGAAATCAGGAGTTCGACAGCAGCAATGCAAGCGTCTGTGATCATCTG 300
Db 8296 CAACCCCAAGAAATCAGGAGTTCGACAGCAGCAATGCAAGCGTCTGTGATCATCTG 8237
QY 301 CCGAAGAGCTTGGCTCAATCAAGCGGACGTGAAGCAATCAAAAGGAGGAATAGTCAA 360
Db 8236 CCGAAGAGCTTGGCTCAATCAAGCGGACGTGAAGCAATCAAAAGGAGGAATAGTCAA 8177
QY 361 AGCAGACGGCGGGCGGGCGGCGGAGCAGCAGCAGCAGCAGGAGTGGGGCTCTG 420
Db 8176 AACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8117
QY 421 CCAGGTACCGGGCGGGCGGCGGAGGTCGCCAGGTGCCAGGTTCCCGGAGGCGCCCTCTCC 480
Db 8116 CCAATACCGGACGAAACCAACAGCAATATCCCAATTCGCGAAACCCCTCTCTCC 8057
QY 481 CTGAGTGGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 8056 CTAATATAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 7997
QY 541 GCGGGCAGCAACTAAGAGATGACGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 600
Db 7996 ACGACACAAACTAAGAGATGACGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 7937
QY 601 GTGGGGTCNCCCTGGCACACGCGGGGTCCTCCAGCCGCCCGGAGACCCCGGAGA 656
Db 7936 ATAAATCGCCTTAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 7881
RESULT 9
AR005195
LOCUS AR005195 1425 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5747645.
ACCESSION AR005195
VERSION AR005195.1 GI:3966074
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and cytoplasmic antiproteinase-3 and coding sequences
JOURNAL Patent: US 5747645-A 1 05-MAY-1998;
FEATURES Location/Qualifiers
source 1. 1425
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
RESULT 10
I81228
LOCUS I81228 1425 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5710026.
ACCESSION I81228
VERSION I81228.1 GI:3209518
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and cytoplasmic antiproteinase-3 and coding sequences
JOURNAL Patent: US 5710026-A 1 20-JAN-1998;
FEATURES Location/Qualifiers
source 1. 1425
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
RESULT 11
I82208
LOCUS I82208 1425 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5712117.
ACCESSION I82208
VERSION I82208.1 GI:3210505
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and coding sequences
JOURNAL Patent: US 5712117-A 1 27-JAN-1998;
FEATURES Location/Qualifiers
source 1. 1425
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
RESULT 12
I82209
LOCUS I82209 1425 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5712117.
ACCESSION I82209
VERSION I82209.1 GI:3210506
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1425)
AUTHORS Sprecher, C.A.
TITLE Cytoplasmic antiproteinase-2 and coding sequences
JOURNAL Patent: US 5712117-A 1 27-JAN-1998;
FEATURES Location/Qualifiers
source 1. 1425
BASE COUNT 437 a 301 c 357 g 330 t
ORIGIN
Query Match 11.2%; Score 73.4; DB 6; Length 1425;
Best Local Similarity 98.7%; Pred. No. 3.2e-07;
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 333 GAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGG 392
Db 7 GGAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGG 66
QY 393 CAGCAGCAGCAGGAG 407
Db 67 CAGCAGCAGCAGGAG 81
```



\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1049: contig of 1049 bp in length  
 1050 1149: gap of unknown length  
 1150 2399: contig of 1150 bp in length  
 2300 2399: gap of unknown length  
 2400 4013: contig of 1614 bp in length  
 4014 4113: gap of unknown length  
 4114 5129: contig of 1016 bp in length  
 5130 5229: gap of unknown length  
 5230 6293: contig of 1064 bp in length  
 6294 6393: gap of unknown length  
 6394 7511: contig of 1118 bp in length  
 7512 7612: gap of unknown length  
 7613 8744: contig of 1132 bp in length  
 8745 8844: gap of unknown length  
 8845 10007: contig of 1164 bp in length  
 10008 10107: gap of unknown length  
 10108 11279: contig of 1172 bp in length  
 11280 11379: gap of unknown length  
 11380 12440: contig of 1061 bp in length  
 12441 12540: gap of unknown length  
 12541 13891: contig of 1351 bp in length  
 13892 13991: gap of unknown length  
 13992 15136: contig of 1145 bp in length  
 15137 15236: gap of unknown length  
 15237 16597: contig of 1261 bp in length  
 16598 17707: contig of 1110 bp in length  
 17708 17807: gap of unknown length  
 17808 19233: contig of 1426 bp in length  
 19234 19333: gap of unknown length  
 19334 20707: contig of 1374 bp in length  
 20708 21907: contig of 1100 bp in length  
 21908 22007: gap of unknown length  
 22008 23207: contig of 1200 bp in length  
 23208 23307: gap of unknown length  
 23308 24909: contig of 1602 bp in length  
 24910 25009: gap of unknown length  
 25010 26420: contig of 1411 bp in length  
 26421 26520: gap of unknown length  
 26521 28020: contig of 1500 bp in length  
 28021 28120: gap of unknown length  
 28121 29722: contig of 1602 bp in length  
 29723 29822: gap of unknown length  
 29823 31076: contig of 1254 bp in length  
 31077 31176: gap of unknown length  
 31177 32290: contig of 1114 bp in length  
 32291 32390: gap of unknown length  
 32391 33610: contig of 1220 bp in length  
 33611 33710: gap of unknown length  
 33711 34987: contig of 1277 bp in length  
 34988 35087: gap of unknown length  
 35088 36268: contig of 1181 bp in length  
 36269 36368: gap of unknown length  
 36369 38071: contig of 1703 bp in length  
 38072 38171: gap of unknown length  
 38172 39859: contig of 1688 bp in length  
 39860 39959: gap of unknown length  
 39960 41056: contig of 1097 bp in length  
 41057 41156: gap of unknown length  
 41157 42335: contig of 1179 bp in length  
 42336 42435: gap of unknown length  
 42436 44405: contig of 1970 bp in length  
 44406 44505: gap of unknown length  
 44506 46527: contig of 2022 bp in length  
 46528 46627: gap of unknown length  
 46628 48152: contig of 1525 bp in length

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
 Bowley, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buick, J., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, F., Carter, K., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Dean, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Honsi, F., Howard, S., Huber, J., Huljk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, J., Johnson, K., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegad, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sison, I.,  
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

2 (bases 1 to 113684)  
 Worley, K.C.

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 113684)  
 Worley, K.C.

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

On Jul 11, 2002 this sequence version replaced gi:17974427.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GIYQ

Center clone name: CH230-197M15

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 46493 bases at least Q40

Consensus quality: 49763 bases at least Q30

Consensus quality: 51555 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 65 contigs. The true order of the pieces

SOURCE  
ORGANISM  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 65711)  
AUTHORS  
Birren,B., Nusbaum,C. and Lander,E.  
TITLE  
Mus musculus, clone RP23-133N21  
REFERENCE  
Unpublished  
2 (bases 1 to 65711)  
AUTHORS  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., DeAlarano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kellis,C., LaRoque,K., Lamazares,R.,  
Lander,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,D.,  
Mihova,T., Miengva,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE  
Direct Submission  
Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 65711)  
REFERENCE  
AUTHORS  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., DeAlarano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Miengva,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.  
TITLE  
Direct Submission  
Submitted (13-AGU-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 13, 2002 this sequence version replaced gi:20128326.  
COMMENT  
All repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Web site: <http://www-seq.wi.mit.edu>  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>

```

Project name: 122422
Center project name: 122422
Center clone name: 133_N_21
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
*

```



Fri Jan 10 08:43:24 2003

us-09-581-500b-12.rge

Page 15

Search completed: January 10, 2003, 00:16:31  
Job time : 2751.88 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:55:35 ; Search time 311.247 Seconds  
(without alignments)  
4746.432 Million cell updates/sec

Title: US-09-581-500B-12  
Perfect score: 656  
Sequence: 1 gccacaacaaatgaat.....cccaccgcgagaccgcgga 656

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_101002.*			
1:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*		
2:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*		
3:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*		
4:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*		
5:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*		
6:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*		
7:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*		
8:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*		
9:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*		
10:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*		
11:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*		
12:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*		
13:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*		
14:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*		
15:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*		
16:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*		
17:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*		
18:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*		
19:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*		
20:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*		
21:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*		
22:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*		
23:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*		
24:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	652.6	99.5 20 AAX88553	Human chromosome 1
2	441.6	67.3 13202 24 ABL33484	Human immune syste
3	367.8	56.1 13202 24 ABL33485	Human immune syste
4	100	15.2 186 21 AAC13994	Human secreted pro
5	73.4	11.2 1425 17 AAT35220	Cytoplasmic antipr
6	73.4	11.2 1425 21 AAZ39749	Human cytoplasmic
7	73	11.1 1325 24 ABK83894	Human cDNA differe
8	50.4	7.7 10732 21 AAL10594	Gene encoding a su
9	50.2	7.7 795 19 AAV55830	FLGA insert stabil

c	10	49.2	7.5	1925	20	AAX90924	Estein Barr Virus
	11	47.2	7.2	3103	23	AAS85091	DNA encoding novel
	12	46.8	7.1	5222	21	AAA5964	Human G713 5'-regu
	13	46.8	7.1	5566	21	AAQ55967	Human G713 encodin
	14	46.4	7.1	286	16	AAQ95180	Simple tandem repe
	15	46.4	7.1	1017	20	AAX17058	Human gene expres
	16	46	7.0	916	22	AAF80065	Amino terminus of
	17	46	7.0	3563	20	AAX23429	Human SCA6 DNA fra
	18	46	7.0	3596	19	AAV61588	Alpha-1A calcium c
	19	46	7.0	3632	19	AAV61586	Alpha-1A calcium c
c	20	45.6	7.0	599	22	ABA60694	Human foetal liver
	21	45.6	7.0	599	22	ABA63254	Human foetal liver
c	22	45.6	7.0	599	22	ABA28778	Probe #7244 for ge
	23	45.6	7.0	599	22	ABA30484	Probe #950 for ge
c	24	45.6	7.0	599	22	AAK08976	Human brain expres
	25	45.6	7.0	599	22	AAK11722	Human brain expres
c	26	45.6	7.0	599	22	AAK34867	Human bone marrow
	27	45.6	7.0	599	22	AAK37465	Human bone marrow
c	28	45.6	7.0	599	22	AAI16860	Probe #6793 for ge
	29	45.6	7.0	599	22	AAI18270	Probe #9203 for ge
c	30	45.6	7.0	599	22	AAI40583	Probe #9269 used t
	31	45.6	7.0	599	22	AAI43326	Probe #12012 used
c	32	45.6	7.0	599	24	ABS09398	Human genome-deriv
	33	45.6	7.0	599	24	ABS11444	Human genome-deriv
c	34	45.6	7.0	840	24	ABQ35494	Oligonucleotide fo
	35	45.6	7.0	840	24	ABQ35495	Oligonucleotide fo
c	36	45.6	7.0	855	22	AAH03166	Human cDNA clone (
	37	45.6	7.0	1881	22	AAH13633	Human cDNA sequenc
c	38	45.6	7.0	2027	24	ABN59664	Novel human coding
	39	45.6	7.0	6545	23	ABL14373	Drosophila melanog
c	40	45.6	7.0	10438	23	ABL14372	Drosophila melanog
	41	45.4	6.9	799	19	AAV55831	Nucleotide sequenc
	42	45.4	6.9	1926	21	AAA50254	Epstein Barr virus
	43	45.4	6.9	1926	22	AAF82902	EBV tethering prot
	44	45.4	6.9	2493	23	AAS88122	DNA encoding novel
	45	45.4	6.9	2580	21	AAV75454	Nucleotide sequenc

ALIGNMENTS

RESULT 1  
AAX88553 AAX88553 standard; DNA: 656 BP.  
XX AAX88553;  
XX  
DT 10-SEP-1999 (first entry)  
XX  
DE Human chromosome 18q YAC clone nucleotide sequence #12.  
XX  
KW Human chromosome 18q; mood disorder; polymorphic marker; detection;  
KW identification; trinucleotide repeat expansion; schizophrenia;  
KW anxiety disorder; adjustment disorder; personality disorder;  
KW nucleotide triplet repeat; ss.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO9932643-A2.  
XX  
PD 01-JUL-1999.  
XX  
PF 17-DEC-1998; 98WO-EP08543.  
XX  
PR 18-DEC-1997; 97GB-0026804.  
XX  
PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
XX  
PI Del-Favero J, Raeynaekers P, Van Broeckhoven C;  
XX WPI; 1999-418934/35.  
XX







KW	gene therapy; chromosome mapping; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP1033401-A2.
XX	
PD	06-SEP-2000.
XX	
PF	21-FEB-2000; 2000EP-0200610.
XX	
PR	26-FEB-1999; 99US-0122487.
XX	
PA	(GEST ) GENSET.
PI	Dumas Milne Edwards J, Duclert A, Giordano J;
XX	
DR	WPI; 2000-500381/45.
XX	
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX	
XX	Claim 1; SEQ ID 18069; 71pp + CD-ROM; English.
XX	
CC	The present sequence is one of a large number of 5' ESTs derived from
CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC	identified within the present sequence. The 5' ESTs were prepared from
CC	total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences
CC	derived from the 5' ends of mRNAs and even in those cases where longer
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC	They are used to obtain upstream regulatory sequences and to design
CC	expression and secretion vectors.
XX	
SQ	Sequence 186 BP; 52 A; 43 C; 57 G; 34 T; 0 other;
Query Match	15.2%; Score 100; DB 21; Length 186;
Best Local Similarity	99.1%; Pred. No. 9e-16; Indels 1; Gaps 1;
Matches 111; Conservative	0; Mismatches
Qy	296 ATCTGCGGAAGCCCTTCCTCAATCAAGCGGCGACGTGAAGCATCTACAAAGGAGGAATA 355 
Db	1 ATCTGCGGAAGCCCTTCCTCAATCAAGCGGCGACGTGAAGCATCTACAAAGGAGGAATA 60 
Qy	356 GTCAAAAGCAGCAGCGGGCGGGCGGCGGCGCAGCAGCAGCAGCAGGAG 407 
Db	61 GTCAAAAGCAGCAGCGGGCGGGCGGCGGCGGCGAG-AGCAGCAGCAGCAGGAG 111 
RESULT 5	
AAAT35220	
ID	AAAT35220 standard; cDNA; 1425 BP.
XX	
AC	AAAT35220;
XX	
DT	05-DEC-1996 (first entry)
XX	
DE	Cytoplasmic antiproteinase-2 protein cDNA.
XX	
KW	Cytoplasmic antiproteinase-2 protein; CAP-2; serpin;
KW	serine protease inhibitor; antiinflammatory; apoptosis; ss.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS 92..1216
XX	/*tag= a

Sequence 1425 BP; 437 A; 301 C; 357 G; 330 T; 0 other;  
AA

\_\_\_\_\_

.....



CC	rapid and reliable method of stably expressing multiple genes in transfected cells. The episomes are useful in the transfection of genes encoding receptors, transporter proteins, ion channels, adhesion molecules and transcription factors. The episomes carrying desired genes can also be used to transfect cells in gene therapy, antisense therapy, for gene amplification, cell immortalisation, etc.
XX	
SQ	Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;
Query Match	7.5%; Score 49.2; DB 20; Length 1925;
Best Local Similarity	49.8%; Pred. No. 0.014;
Matches 120; Conservative 1;	Mismatches 120; Indels 0; Gaps
QY	346 AGGAGGAATAGTCAAAAGCAGACAGCGGGCGCGCGGCAGCACAGCAGCAGCAGG 405 
Db	705 AGGAGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 764
QY	406 AGGTGGGGGCTCTGCCAGGTACC GGCGGGGCAGGCAGCAGGAGTGCCCCAGGTTCCCGG 465 
Db	765 AGGAGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCA 824
QY	466 GAGGCCACTCTTCCTGGAGTGGCGTAGAGAGGGGAAGGAGGAGGCCAGACAGCGAA 525
Db	825 GGAGGGCCAGGACAGGAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCA 884
QY	526 TCAGAGCGCAGCAAAAGCGGGGCAGAACTANGAATAATGACSGCGGAGGCGGCGCGGAAA 595 
Db	885 GGAGCAGGAGGGGCAGGAGGGGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGA 944
QY	586 G 586
Db	945 G 945
RESULT 11	
AAS85091/C	
ID	AAS85091 standard; cDNA; 3103 BP.
XX	
AC	AAS85091;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #20895.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG20904.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
PT	
XX	
PS	Claim 1; SEQ ID No 20895; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AA564197-AA594564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3103 BP; 669 A; 871 C; 899 G; 664 T; 0 other;

Query Match 7.2%; Score 47.2; DB 23; Length 3103;  
 Best Local Similarity 54.7%; Pred. No. 0.052;  
 Matches 94; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
 QY 346 AGGAGGATATGTAAGAGCAGCAGCGCGCGCGCGAGCAGCAGCAGCAGG 405  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 966 AGGGGAGGAGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 907  
 QY 406 AGGTGGGGGGCTCTGCCAGGTACCGCGCGCGCGCGAGGTGCCAGGTTCCCGCG 465  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 906 AGGCAGCTCCCTGGCTTCGCGCGCGCGCTCTGCCCGAATGCGTGGCATCTGTTTCGAG 847  
 QY 466 GAGGCGACCTCTTCCCTGGAGTGGTGAGAGAGGGGAGGAGGAGGCCAG 517  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 846 TAGCCCCCATAGACGAGCAGATGCCCGCTGGGGAGTGACGACATCTGGCAG 795

RESULT 12  
 AA55964  
 ID AAA55964 standard; DNA; 5222 BP.  
 XX  
 AC AAA55964;  
 XX  
 DT 05-SEP-2000 (first entry)  
 XX  
 DE Human G713 5'-regulatory region, exon 1 and 5'-end of intron 1.  
 XX  
 KW Human; chromosome 13; G713; chromosome 13q31-q33; schizophrenia;  
 KW biallelic marker; polymorphism; central nervous disease; detection;  
 KW neuroleptic; G713 gene expression inhibitor; genotyping;  
 KW brain disorder; psychiatric disorder; bipolar disorder; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200022122-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 12-OCT-1999; 99WO-IB01730.  
 XX  
 PR 13-OCT-1998; 98US-0103955.  
 PR 30-OCT-1998; 98US-0106457.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D, Essiloux L;  
 XX WPI; 2000-317979/27.  
 XX  
 PT Novel polynucleotide of human G713 gene useful for diagnosis and

PT prophylactic treatment of brain, psychiatric disorders like  
 PT schizophrenia and bipolar disorders -  
 XX  
 XX Claim 1; Page 211-212; 271pp; English.  
 XX  
 CC The present invention describes an isolated, purified or recombinant  
 CC polynucleotide (PN) (I) comprising a contiguous span of 8 to 50  
 CC nucleotides, where the span includes a G713 or chromosome 13q31-q33  
 CC related biallelic marker. (I) has neuroleptic activity and can be used  
 CC as a G713 gene expression inhibitor. (I) can be used genotyping to  
 CC estimate the frequency of an allele of a G713 or chromosome 13q31-q33  
 CC related biallelic marker in a population, and of a haplotype for a set  
 CC of biallelic markers in a population. (I) is also useful in detecting  
 CC an association between a haplotype and a trait. The frequency is used  
 CC for detecting an association between a genotype and a trait being  
 CC schizophrenia. The genotype is used to determine whether an individual  
 CC is at risk of developing schizophrenia. (I) can also be used as a  
 CC medicament against several disorders preferably brain, psychiatric  
 CC disorders such as schizophrenia and bipolar disorder. Early  
 CC identification of risk of developing schizophrenia is possible, which  
 CC would enable early and/or prophylactic treatment. AAA55964 to AAA55966  
 CC represent human G713 genomic DNA sequences; AAA55967 encodes the human  
 CC G713 protein AA590962; AAA55968 encodes the murine G713 protein  
 CC AA590963; AAA55992 to AAA56030 represent human chromosome 13q31-q33 locus  
 CC biallelic markers A12 to A49; AAA55969 to AAA55991, and AAA56031 and  
 CC AAA56032 represent PCR primers used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 5222 BP; 1278 A; 1304 C; 1439 G; 1176 T; 25 other;  
 Query Match 7.1%; Score 46.8; DB 21; Length 5222;  
 Best Local Similarity 58.7%; Pred. No. 0.078;  
 Matches 81; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 346 AGGAGGATATGTAAGAGCAGCAGCGCGCGCGCGAGCAGCAGCAGCAGG 405  
 Db AGC 4023  
 QY 406 AGGTGGGGGGCTCTGCCAGGTACCGCGCGCGCGAGCAGCAGCAGGTTGCCAGGTTCCCGCG 465  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 4024 AGGAGCCCTCTGTGCCCGCGCTCTGTGGCGAGCATGGGGAGTCTCTGCCCGCCCGCCAGG 4083  
 QY 466 GAGGCGACCTCTTCCCTGG 483  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 4084 CACACAGACTCTCTCCG 4101  
 RESULT 13  
 AA55967  
 ID AAA55967 standard; cDNA; 5566 BP.  
 XX  
 AC AAA55967;  
 XX  
 DT 05-SEP-2000 (first entry)  
 XX  
 DE Human G713 encoding cDNA SEQ ID NO:4.  
 XX  
 KW Human; chromosome 13; G713; chromosome 13q31-q33; schizophrenia;  
 KW biallelic marker; polymorphism; central nervous disease; detection;  
 KW neuroleptic; G713 gene expression inhibitor; genotyping;  
 KW brain disorder; psychiatric disorder; bipolar disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200022122-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 12-OCT-1999; 99WO-IB01730.  
 XX  
 PR 13-OCT-1998; 98US-0103955.  
 PR 30-OCT-1998; 98US-0106457.  
 XX

[illegible]



Search completed: January 9, 2003, 21:33:45  
Job time : 327.247 secs

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:50:31 ; Search time 55.2338 Seconds  
(without alignments)  
3642.332 Million cell updates/sec

Title: US-09-581-500B-12  
Perfect score: 656  
Sequence: 1 gccacaacaaatgaat.....cccacgcgagaccgcgcga 656

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.4	11.2	1425	1 US-08-464-148-1	Sequence 1, Appli
2	73.4	11.2	1425	1 US-08-385-500-1	Sequence 1, Appli
3	73.4	11.2	1425	1 US-08-846-784-1	Sequence 1, Appli
c 4	72	11.0	7218	1 US-08-232-463-14	Sequence 14, Appli
c 5	49.4	7.5	1104	4 US-09-009-816-1	Sequence 1, Appli
6	46.4	7.1	286	2 US-08-332-766A-6	Sequence 6, Appli
7	46	7.0	3563	4 US-09-041-886-20	Sequence 20, Appli
8	46	7.0	3596	2 US-08-779-801-5	Sequence 5, Appli
9	46	7.0	3596	2 US-08-238-441-5	Sequence 5, Appli
10	46	7.0	3632	2 US-08-779-801-3	Sequence 3, Appli
11	46	7.0	3632	4 US-09-238-441-3	Sequence 3, Appli
12	45.8	7.0	325	2 US-08-531-927B-3	Sequence 3, Appli
13	45.4	6.9	1926	4 US-09-249-585A-2	Sequence 2, Appli
14	45.4	6.9	2580	3 US-09-050-863-2	Sequence 2, Appli
15	45.4	6.9	2580	4 US-09-359-081-2	Sequence 2, Appli
c 16	45.4	6.9	5452	2 US-09-130-114-1	Sequence 1, Appli
17	45.4	6.9	9600	4 US-08-910-647-1	Sequence 1, Appli
18	45.4	6.9	9600	4 US-09-620-925-1	Sequence 1, Appli
19	45.4	6.9	10596	1 US-07-884-811-15	Sequence 15, Appli
20	45.4	6.9	10596	1 US-07-885-971-15	Sequence 15, Appli
21	45.4	6.9	10596	1 US-08-087-783A-15	Sequence 15, Appli
22	45.4	6.9	10596	1 US-08-194-088B-15	Sequence 15, Appli
23	45.4	6.9	10596	2 US-08-194-087-15	Sequence 15, Appli
24	45.4	6.9	10596	5 PCT-US93-04648-15	Sequence 15, Appli
25	44.6	6.8	1028	4 US-08-118-200-1	Sequence 1, Appli
26	44.6	6.8	1028	4 US-08-458-745-1	Sequence 1, Appli
27	44.4	6.8	1488	4 US-08-406-030A-3	Sequence 3, Appli

28	44.2	6.7	2056	4 US-09-334-601-12	Sequence 12, Appli
29	43.6	6.6	4362	2 US-08-455-073A-1	Sequence 1, Appli
30	43.6	6.6	7808	2 US-08-149-097D-22	Sequence 22, Appli
31	43.6	6.6	7808	3 US-08-949-386-22	Sequence 22, Appli
32	43.6	6.6	7808	3 US-08-450-562-22	Sequence 22, Appli
33	43.6	6.6	7808	4 US-08-984-709A-22	Sequence 22, Appli
34	43.6	6.6	7808	4 US-08-450-272-22	Sequence 22, Appli
c 35	43	6.6	670	4 US-09-009-816-3	Sequence 3, Appli
36	42.6	6.5	300	4 US-09-135-994-3	Sequence 1, Appli
37	42.2	6.4	477	4 US-09-135-994-1	Sequence 1, Appli
38	42.2	6.4	2115	2 US-08-474-379C-60	Sequence 60, Appli
39	42.2	6.4	2115	3 US-09-146-249A-60	Sequence 60, Appli
40	42.2	6.4	2115	3 US-08-206-188B-60	Sequence 60, Appli
41	42.2	6.4	2770	4 US-09-008-697A-13	Sequence 13, Appli
c 42	42.2	6.4	4897	6 5196516-7	Patent No. 5196516
43	42	6.4	9972	3 US-08-836-022A-3	Sequence 3, Appli
c 44	42	6.4	9972	4 US-09-427-048A-3	Sequence 3, Appli
c 45	42	6.4	53526	3 US-08-658-136-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-464-148-1  
; Sequence 1, Application US/08464148  
; Patent No. 5710026  
; GENERAL INFORMATION:  
; APPLICANT: Sprecher, Cindy A.  
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-2 AND CODING SEQUENCES  
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourile and Crew  
; STREET: Stewart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/464,148  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/385,500  
; FILING DATE: 08-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1425 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 92..1213  
; OTHER INFORMATION: /product= "CYTOPLASMIC  
; OTHER INFORMATION: ANTIPROTEINASE-2 PROTEIN"  
US-08-464-148-1

Query Match 11.2%; Score 73.4; DB 1; Length 1425;

Best Local Similarity 98.7%; Pred. No. 1.6e-10;  
Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 333 GAAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCGCGCGCAG 392

Db 7 GGAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCGCGCAG 66

QY 393 CAGCAGCAGCAGGAG 407

Db 67 CAGCAGCAGCAGGAG 81

## RESULT 2

US-08-385-500-1  
; Sequence 1, Application US/08385500  
; Patent No. 5712117  
; GENERAL INFORMATION:  
; APPLICANT: Sprecher, Cindy A.  
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-2 AND  
; TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: US  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/385,500  
; FILING DATE: 08-FEB-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/385,500  
; FILING DATE: 08-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W.  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 467-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1425 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 92..1213  
; OTHER INFORMATION: /product= "CYTOPLASMIC  
; OTHER INFORMATION: ANTIPROTEINASE-2 PROTEIN"  
US-08-385-500-1

Query Match 11.2%; Score 73.4; DB 1; Length 1425;

Best Local Similarity 98.7%; Pred. No. 1.6e-10;

Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 333 GAAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCGCAG 392

Db 7 GGAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCGCAG 66

QY 393 CAGCAGCAGCAGGAG 407

Db 67 CAGCAGCAGCAGGAG 81

## RESULT 3

US-08-846-784-1

Sequence 1, Application US/08846784

Patent No. 5747645

GENERAL INFORMATION:

APPLICANT: Sprecher, Cindy A.

TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-2 AND

TITLE OF INVENTION: CYTOPLASMIC ANTIPROTEINASE-3 AND CODING SEQUENCES

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend Kourie and Crew

STREET: Steuart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,784

FILING DATE: 30-APR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/385,500

FILING DATE: 08-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W.

REGISTRATION NUMBER: 31,990

REFERENCE/DOCKET NUMBER: 13952-21

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 467-9600

TELEFAX: (415) 543-5043

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1425 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 92..1213

OTHER INFORMATION: /product= "CYTOPLASMIC

OTHER INFORMATION: ANTIPROTEINASE-2 PROTEIN"

US-08-846-784-1

Query Match 11.2%; Score 73.4; DB 1; Length 1425;

Best Local Similarity 98.7%; Pred. No. 1.6e-10;

Matches 74; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 333 GAAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCAG 392

Db 7 GGAGCATCTACAAAGAGGAGTAATGTCAAAGCAGCAGCGCGCGCGCGCGCAG 66

QY 393 CAGCAGCAGCAGGAG 407

Db 67 CAGCAGCAGCAGGAG 81

## RESULT 4

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.

; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

; NUMBER OF SEQUENCES: 52

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 1800 Diagonal Road, Suite 500

;; CITY: Alexandria  
;; STATE: VA  
;; COUNTRY: USA  
;; ZIP: 22313-0299  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patentin Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/232,463  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/935,313  
;; FILING DATE:  
;; APPLICATION NUMBER: EP 91 114 300.6  
;; FILING DATE: 26-AUG-1991  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BENT, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (703)836-9300  
;; TELEFAX: (703)683-4109  
;; TELEX: 899149  
;; INFORMATION FOR SEQ ID NO: 14:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 7218 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; IMMEDIATE SOURCE:  
;; CLONE: pTZ5pt-Fls  
;; US-08-232-463-14

Query Match 11.0%; Score 72; DB 1; Length 7218;  
Best Local Similarity 3.8%; Pred. No. 7.6e-10;  
Matches 15; Conservative 234; Mismatches 142; Indels 0; Gaps 0;  
QY 248 AAAGATCAGGATTCGACAGCAGCAATCGAACGGTCTGAGTATCTGCGGGAAG 307  
Db 1425 RRR 1366  
QY 308 CTTGCGCTCAATCAGCGGACGTGAAGCATCTACAAAGAGGAATGTCACAGCAGCA 367  
Db 1365 RRR 1306  
QY 368 GCGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 427  
Db 1305 RRR 1246  
QY 428 CGGCGCGGCGCAGCAGGAGTGCAGGTTCCCGGAGGCCACTTCCCTGGAGT 487  
Db 1245 RRR 1186  
QY 488 GCGTCAGAGAGGGAAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 547  
Db 1185 RRR 1126  
QY 548 AGGAATCANGAATGACSCGGGAGCGCGCGGGAAGAAANTCCTGGGCGTGTGGGG 607  
Db 1125 RRR 1066  
QY 608 TCNCCCTGGCAGCAGCGGGGTCCCAAGCCC 638  
Db 1065 TCGCAAGCTCCCTCGACCTCAGCAGCAAGCTC 1035

RESULT 5  
US-09-009-816-1/c  
; Sequence 1, Application US/09009816  
; Patent No. 6436667

;; GENERAL INFORMATION:  
;; APPLICANT: German, Michael  
;; APPLICANT: Permutt, M. Alan  
;; APPLICANT: Inoue, Hiroshi  
;; TITLE OF INVENTION: Human Nkx-6.1 Polypeptide-Encoding  
;; NUCLEOTIDE SEQUENCES: 21  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Bozicevic & Reed, LLP  
;; STREET: 285 Hamilton Ave, Suite 200  
;; CITY: Palo Alto  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94301  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FastSeq for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/009,816  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Francis, Carol L.  
;; REGISTRATION NUMBER: 36,513  
;; REFERENCE/DOCKET NUMBER: 9076/082CIP2  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 650-327-3400  
;; TELEFAX: 650-327-3231  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1104 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; FEATURE:  
;; NAME/KEY: Coding Sequence  
;; LOCATION: 1..1101  
;; OTHER INFORMATION:  
;; US-09-009-816-1  
Query Match 7.5%; Score 49.4; DB 4; Length 1104;  
Best Local Similarity 50.9%; Pred. No. 0.00041;  
Matches 113; Conservative 1; Mismatches 108; Indels 0; Gaps 0;  
QY 361 AGCAGCAGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420  
Db 426 ACCCG 367  
QY 421 CCAGGTACCG 480  
Db 366 ACCCGAACCAGCG 307  
QY 481 CTGCGAGTCGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
Db 306 CAGGATATTGTGATGCGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGGGTGTGGGT 247  
QY 541 GCGCGCGCAGGAACCTANGAGAAATGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 582  
Db 246 GATGAGAGCCCGCCCGTGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205  
RESULT 6  
US-08-332-766A-6  
; Sequence 6, Application US/08332766A  
; Patent No. 5843647  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREYS, Alec J.

```

RESULT 7
US-93-041-886-20
Sequence 20, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Sharroo
TITLE OF INVENTION: Proapoptotic
TITLE OF INVENTION: Polypeptides
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores L
STREET: 4370 La Jolla Village
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

RESULT 8
US-08-779-801-5
; Sequence 5, Application US/08779801
; Patent NO. 5853995
; GENERAL INFORMATION:
; APPLICANT: Lee, Cheng-Chi
; TITLE OF INVENTION: Large Scale Genotyping of
; TITLE OF INVENTION: Diseases and a Diagnostic
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Benjamin Aaron Adler, Ph.D., J.D.
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/779,801
; FILING DATE: January 7, 1997

```



[illegible]

```

Db 3162 AGGATGTGAGAGCGGGTCCAGGCCGCCGCCGAGGAGTCCCCCAGGGGCTGTGACAC 3221
QY 585 AGAAANTCTCGGGGCTGTGGGGTGNCCCTGGCACCAGCGCGGGTGTCCCAAGCCCCACCG 643
Db 3222 GCGGGGCCCGGGTGGCGGCACTGTGGCCGCGCAGCTGTCCAGGGGCCCGCGGGTCCCCG 3280

RESULT 12
US-08-531-927B-3
; Sequence 3, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/531.927B
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP H6-251600
; FILING DATE: 21-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Grabau, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: ATH95-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 145
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: 194
US-08-531-927B-3

Query Match 7.0%; Score 45.8; DB 2; Length 325;
Best local similarity 69.4%; Pred. No. 0.0023;
Matches 59; Conservative 2; Mismatches 24; Indels 0; Gaps 0;

QY 332 TGAAGCATCTCAAGAGGAGGAATAGTCAAGCAGCAGCGCGCGCGCGGCGGACGA 391
Db 115 TGAATGTTTCAGCAGCAGCAGCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 174
QY 392 GCACGACGACGACGAGGAGTGGGGCC 416
Db 175 GCACGACGACGACGACGAGSGGNC 199

RESULT 13
US-09-249-585A-2

```



```

; REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match          6.9%; Score 45.4; DB 4; Length 2580;
Best Local Similarity 49.0%; Pred. No. 0.0071;
Matches 142; Conservative 1; Mismatches 145; Indels 2; Gaps 1

QY  346 AGGAGGATAGTCAAGCAGCAGCGCGCGCGCGGCGGCGGCGAGCAGCAGCAGCAGG 405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   896 AGGAGGAGGGCAGGAGGAGGAGGGGCGAGGAGGGGCGCAGGAGCAGGAGGGGCGAGG 955
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  406 AGGTTGGGGCGCTCTGCCAGGTACCCGGCGGGCGGCGAGGCACGGAGGTGCCCGAGGTTCCCGCG 465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   956 AGGGCAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGCAGGAGGAGGGCGCAGG 1015
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  466 GAGGCCACTCTTCCCTTGAGTGTGCTGACAGAGGGGAAGGGAGGAGGCGCAGCAGAGGAA 525
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1016 AGGGGCGAGGA--GCAGGAGGGCGAGCAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCA 1073
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  526 TCAGAGCCAGCAGCAAGCGGCGAGCACTANGAGAAATGACGGGGGAGGCGCGCGGAAA 585
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1074 GGAGCAGGAGGGGCGAGCAGCAGGAGGGGCGAGGAGCGAGGAGGGGCGAGGAGGAGGGGCA 1133
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  586 GAAATCTCTGGGGGTGTGGGGGTCTCCCTTGGCAGCCAGCCCGGGGTGCCCAAG 635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1134 GGAGGGGCGAGCAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 1183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: January 10, 2003, 06:37:32
Job time : 72.2338 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:01:36 ; Search time 85.8611 Seconds  
(without alignments)  
3362.497 Million cell updates/sec

Title: US-09-581-500b-12

Perfect score: 656

Sequence: 1 gccacaacaaataatgaat.....cccaccgcgagaccgcgcga 656

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_NA.\*  
1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2.6/ptodata/2/pubpna/ECT\_NEW\_PUB.seq.\*  
3: /cgn2.6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2.6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2.6/ptodata/2/pubpna/ECTUS\_PUBCOMB.seq.\*  
7: /cgn2.6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2.6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2.6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2.6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2.6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2.6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	91.8	14.0	1476	9	US-09-974-298-94
2	47.8	7.3	2724	12	US-10-044-090-309
3	46.8	7.1	5222	10	US-09-416-384A-1
4	46.8	7.1	5566	10	US-09-416-384A-4
5	45.6	7.0	599	10	US-09-864-761-7244
6	45.6	7.0	599	10	US-09-864-761-5950
7	45.4	6.9	3886	9	US-10-124-800-37
8	44.3	6.8	230	10	US-09-864-761-23975
9	44	6.7	210	10	US-09-864-761-22135
10	44	6.7	293	10	US-09-864-761-18923
11	44	6.7	459	10	US-09-864-761-2182
12	44	6.7	464	10	US-09-864-761-5361
13	43	6.6	527	10	US-09-925-297-272
14	42.8	6.5	311	10	US-09-960-352-14967
15	42.6	6.5	2453	12	US-10-005-858-1
16	42.4	6.5	612	10	US-09-878-574-4578
17	42.2	6.4	1310	9	US-09-849-243-13
18	42.2	6.4	3263	9	US-09-849-243-15
19	42.2	6.4	4286	9	US-09-849-243-14

```

c 20 42 1511 9 US-10-078-650-11 Sequence 11, Appl
c 21 42 3641 9 US-10-078-650-11 Sequence 1, Appl
22 41.6 226 10 US-09-728-445-159 Sequence 159, App
23 41.6 2572 10 US-09-925-300-486 Sequence 486, App
24 41.6 3715 10 US-09-880-107-2300 Sequence 2300, Ap
c 25 41.4 6.3 501 10 US-09-864-761-8824 Sequence 8824, Ap
c 26 41.4 6.3 2336 12 US-10-044-090-212 Sequence 212, App
c 27 41.4 6.3 4281 12 US-10-002-600-23 Sequence 23, Appl
c 28 41.2 6.3 122 10 US-09-864-761-25491 Sequence 25491, A
c 29 41.2 6.3 1518 10 US-09-976-165-20 Sequence 20, Appl
c 30 41.2 6.3 2636 10 US-09-976-165-21 Sequence 21, Appl
c 31 41 6.2 441 10 US-09-893-737-41 Sequence 41, Appl
c 32 41 6.2 557 10 US-09-919-580-475 Sequence 475, App
c 33 41 6.2 572 10 US-09-919-580-570 Sequence 570, App
c 34 41 6.2 1064 10 US-09-804-682-29 Sequence 29, Appl
c 35 40.8 6.2 1160 9 US-09-946-807-108 Sequence 108, App
c 36 40.8 6.2 1160 10 US-09-795-668-108 Sequence 108, App
c 37 40.8 6.2 1160 10 US-09-795-668-108 Sequence 108, App
c 38 40.8 6.2 1503841 9 US-09-946-807-1 Sequence 1, Appl
c 39 40.8 6.2 1503841 10 US-09-795-668-1 Sequence 1, Appl
c 40 40.8 6.2 1503841 10 US-09-795-668-1 Sequence 1097, Ap
c 41 40.6 6.2 401 9 US-09-946-807-1097 Sequence 1097, Ap
c 42 40.6 6.2 401 10 US-09-795-668-1097 Sequence 1097, Ap
c 43 40.6 6.2 401 10 US-09-795-668-1097 Sequence 19, Appl
c 44 40.6 6.2 1362 10 US-09-822-830A-19 Sequence 1185, Ap
c 45 40.6 6.2 1636 10 US-09-954-456-1185

```

#### ALIGNMENTS

```

RESULT 1
US-09-974-298-94
; Sequence 94, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 94
; LENGTH: 1476
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 589880CBI
US-09-974-298-94

```

```

Query Match 14.08; Score 91.8; DB 9; Length 1476;
Best Local Similarity 93.28; Pred. No. 4.6e-15;
Matches 96; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 315 CTCATCAAGCGCGACGTCAAGCATCTACAAAGGAGGAGTACAAAGCAGCAGCGCGG 374
Db 1 CTCATCAAGCGCGACGTCAAGCATCTACAAAGGAGGAGTACAAAGCAGCAGCGCGG 60
QY 375 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
Db 61 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103

```

```

RESULT 2
US-10-044-090-309/c
; Sequence 309, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman

```

```

; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 309
; LENGTH: 2724
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 300437.18
; NAME/KEY: unsure
; LOCATION: 2060
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-309

Query Match          7.3%; Score 47.8; DB 12; Length 2724;
Best Local Similarity 53.5%; Pred. No. 0.003;
Matches 100; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 362 GCACGACGCGCGCGCGCGCGCGCGAGCAGCAGCAGCAGGAGTGGGGCTCTGCG 421
Db 517 GCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 458
QY 422 CAGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 481
Db 457 GCACCTCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398
QY 482 TGGAGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
Db 397 GCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
QY 542 GCGGGCGCA 548
Db 337 GGAGGCA 331

RESULT 3
US-09-416-384A-1
; Sequence 1, Application US/09416384A
; Patent No. US20020081584A1
; GENERAL INFORMATION:
; APPLICANT: BLUMENFELD, Marta
; APPLICANT: BOUGUELERET, Lydie
; APPLICANT: CHOMAKOV, Ilya
; APPLICANT: COHEN, Daniel
; APPLICANT: ESSIOUX, Laurent
; TITLE OF INVENTION: Genes, proteins and biallelic markers related to central...
; FILE REFERENCE: GENSET.045AUS
; CURRENT FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: US/09/416,384A
; PRIOR FILING DATE: 1999-10-30
; PRIOR APPLICATION NUMBER: 60/106,457
; PRIOR FILING DATE: 1998-10-12
; PRIOR APPLICATION NUMBER: 60/103,955
; PRIOR FILING DATE: 1998-10-12
; PRIOR APPLICATION NUMBER: 60/132,277
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 5222
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1076..3075
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 3076..4643
; OTHER INFORMATION: exon 1
; NAME/KEY: allele

```

```

; LOCATION: 4872
; OTHER INFORMATION: 8-58-301 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 3606
; OTHER INFORMATION: insertion of AGAG in SEQID4
; NAME/KEY: primer_bind
; LOCATION: 4572..4587
; OTHER INFORMATION: 8-58.pu
; NAME/KEY: primer_bind
; LOCATION: 4990..5005
; OTHER INFORMATION: 8-58.rp complement
; NAME/KEY: misc_binding
; LOCATION: 4849..4895
; OTHER INFORMATION: 8-58-301.probe
; NAME/KEY: primer_bind
; LOCATION: 4853..4871
; OTHER INFORMATION: 8-58-301.mis
; NAME/KEY: primer_bind
; LOCATION: 4873..4891
; OTHER INFORMATION: 8-58-301.mis complement
; NAME/KEY: misc_feature
; LOCATION: 148,686,902,1258,1322,2440,2794,2852,3018..3023,3052..3053,3064
; OTHER INFORMATION: n=a, g, c or t
US-09-416-384A-1

Query Match          7.1%; Score 46.8; DB 10; Length 5222;
Best Local Similarity 58.7%; Pred. No. 0.0074;
Matches 81; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 346 AGGAGGAATAGTCAACAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
Db 3964 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4023
QY 406 AGGTGGGGCGCTCTGCCAGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465
Db 4024 AGGAGCCCTCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4083
QY 466 GAGGCCACCTCTCTCCCTG 483
Db 4084 CACACAGACTCTCTCTCG 4101

RESULT 4
US-09-416-384A-4
; Sequence 4, Application US/09416384A
; Patent No. US20020081584A1
; GENERAL INFORMATION:
; APPLICANT: BLUMENFELD, Marta
; APPLICANT: BOUGUELERET, Lydie
; APPLICANT: CHOMAKOV, Ilya
; APPLICANT: COHEN, Daniel
; APPLICANT: ESSIOUX, Laurent
; TITLE OF INVENTION: Genes, proteins and biallelic markers related to central...
; FILE REFERENCE: GENSET.045AUS
; CURRENT FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: US/09/416,384A
; PRIOR FILING DATE: 1999-10-30
; PRIOR APPLICATION NUMBER: 60/106,457
; PRIOR FILING DATE: 1999-10-30
; PRIOR APPLICATION NUMBER: 60/103,955
; PRIOR FILING DATE: 1998-10-12
; PRIOR APPLICATION NUMBER: 60/132,277
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: Patent.pm
; SEQ ID NO 4
; LENGTH: 5566
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 659..2032
; NAME/KEY: allele
; LOCATION: 4484

```



	Query Match	7.0%;	Score 45.6;	DB 10;	Length 599;
	Best Local Similarity	49.8%;	Pred. No. 0.0061;		
	Matches 114;	Conservative	0; Mismatches 115;	Indels 0;	Gaps 0;
QY	361	AGCAGCACGGCGGCGGCGGCAGCAGCAGCAGCAGCAGGATGGGGCCCTCTG	420		
Db	313	ACACACATATCGCGCGCGCGCGCGCAGCAGCAGCAGCAGCAGCAAAAG	372		
QY	421	CCAGGTACC GGCGGCGGCGAGGCACAGGAGTCCCAGGTTCCCGCGAGGCCACCTCTTC	480		
Db	373	ACCAGGACGAGTTCAGCTGCTCCCTCCCTCCCTCCCTCCGCTCAAGCCAGGAGTGCT	432		
QY	481	CTGGAGTCCGTAGAGAGGGGAAGGAGAGCCAGCAGCAGANTCAGACGCGGCAA	540		
Db	433	TGGGACTCTGGTCTGCGAGCTGGGAAGGCGAGGGAGCTGTGCTCAGCAGAGGCTGGAA	492		

7 AFFILIATION: CHEN, WEIWEI  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842

```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 23975
; LENGTH: 230
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008757.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.4
; OTHER INFORMATION: SWISSPROT HIT: P48678, EVALUE 1.40e+00
; OTHER INFORMATION: NT HIT: M88356.1, EVALUE 1.40e-01
; OTHER INFORMATION: EST_HUMAN HIT: A1016731.1, EVALUE 3.00e-03
US-09-864-761-23975

Query Match      6.8%; Score 44.8; DB 10; Length 230;
Best Local Similarity 51.2%; Pred. No. 0.0066;
Matches 103; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 361 AGCAGCAGCGGCGGGCGGCGGCAGCAGCAGCAGCAGGCTGGGGCGCTCTG 420
Db 224 ACCAGCAGTATCGCGGCGGCGGCGGCAGCAGCAGCAGCAGCAGGAAAG 165
QY 421 CCAGGTACCGGCGGGCAGCAGCAGGAGTGCCTCCCTCCCTCCCTCCCTCC 480
Db 164 ACCAGGAGGAGTTGCAGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 105
QY 481 CTGGAGTGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 104 TGGGACCTCTGGCTGGCACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 45
QY 541 GCGGCGGCAAGCACTANGAGAA 561
Db 44 AACGTGGAATGAGACTAGAA 24

RESULT 9
US-09-864-761-22135

; Sequence 22135, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US 09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22135
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002044.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: P48678, EVALUE 1.30e+00
; OTHER INFORMATION: NT HIT: AF248054.1, EVALUE 5.00e-04
; OTHER INFORMATION: EST_HUMAN HIT: AW296870.1, EVALUE 3.00e-03
US-09-864-761-22135

Query Match      6.7%; Score 44; DB 10; Length 210;
Best Local Similarity 52.1%; Pred. No. 0.01;
Matches 98; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
```

```

;
; FEATURE:
;
; OTHER INFORMATION: MAP TO AC009954.1
;
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
;
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
;
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
;
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
;
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
;
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
;
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45
;
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
;
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
;
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
;
; OTHER INFORMATION: NT HIT: Z22828.1, EVALUAE 9.00e-59
;
; OTHER INFORMATION: SWISSPROT HIT: P53360, EVALUAE 4.00e-15
;
; OTHER INFORMATION: EST_HUMAN HIT: AW409956.1, EVALUAE 1.00e-58
;
; US-09-864-761-18923
;

```

```

Query Match          6.7%; Score 44; DB 10; Length 293;
Best Local Similarity 73.7%; Pred. No. 0.012;
Matches 56; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY  346 AGGAGGATAGTCAAAAGCAGCAGCGCGCGCGCGCGGCGGCGAGCAGCAGCAGCAGG 405
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   158 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 217

QY  406 AGGTGGGGGCGCTTGC 421
      ||| ||| ||| ||| |||
Db   218 AGGCAGTGGCAGCTGC 233

```

RESULT 11  
US-09-864-761-2182  
; Sequence 2182, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aesmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670

```

1 CURRENT REPLICATION NUMBER: US/97/8034, 70
2
3 CURRENT FILING DATE: 2001-05-23
4
5 PRIOR APPLICATION NUMBER: US 60/180,312
6
7 PRIOR FILING DATE: 2000-02-04
8
9 PRIOR APPLICATION NUMBER: US 60/207,456
10
11 PRIOR FILING DATE: 2000-05-26
12
13 PRIOR APPLICATION NUMBER: US 09/632,366
14
15 PRIOR FILING DATE: 2000-08-03
16
17 PRIOR APPLICATION NUMBER: GB 24263. 6
18
19 PRIOR FILING DATE: 2000-10-04
20
21 PRIOR APPLICATION NUMBER: US 60/236,359
22
23 PRIOR FILING DATE: 2000-09-27
24
25 PRIOR APPLICATION NUMBER: PCT/US01/00666
26
27 PRIOR FILING DATE: 2001-01-30
28
29 PRIOR APPLICATION NUMBER: PCT/US01/00667
30
31 PRIOR FILING DATE: 2001-01-30
32
33 PRIOR APPLICATION NUMBER: PCT/US01/00664
34
35 PRIOR FILING DATE: 2001-01-30
36
37 PRIOR APPLICATION NUMBER: PCT/US01/00669
38
39 PRIOR FILING DATE: 2001-01-30
40
41 PRIOR APPLICATION NUMBER: PCT/US01/00665
42

```

```

, GENERAL INFORMATION:
, APPLICANT: Rosen et al.
, TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
, FILE REFERENCE: PA105
, CURRENT APPLICATION NUMBER: US/09/925,297
, CURRENT FILING DATE: 2001-08-10
, PRIOR APPLICATION NUMBER: PCT/US00/05989
, PRIOR FILING DATE: 2000-03-08
, PRIOR APPLICATION NUMBER: 60/124,270
, PRIOR FILING DATE: 1999-03-12
, NUMBER OF SEQ ID NOS: 928
, SOFTWARE: PatentIn Ver. 2.0
, SEQ ID NO. 272

```



```

Db      139  TCAAGGGCGAGAACGGCGCGCGGAAACAAACAGAGCTGGAGGAAACAGNNGCAG 198
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy      542  GGGGCGAGGAA 552
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      199  AAGNGANGGA 209
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

RESULT 15
US-10-005-858-1
; Sequence 1, Application US/10005858
; Patent No. US20020116729A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING NTTP1
; TITLE OF INVENTION: PHOSPHATASE GENE DISRUPTIONS
; FILE REFERENCE: R-690
; CURRENT APPLICATION NUMBER: US/10/005,858
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/251,802
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2453
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-005-858-1

Query Match          6.5%; Score 42.6; DB 12; Length 2453;
Best Local Similarity 59.5%; Pred. No. 0.07; Indels 0; Gaps
Matches 72; Conservative 0; Mismatches 49;

Qy      346  AGAGGAGATAGTCTAAAGCAGCAGCGCGCGCGCGCGCGCGCAGCAGCAGCAGCAGG 405
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      1826  AGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAGT 1885
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy      406  AGGTGGGGGCTCTGCCAGGTACCGGGGGGGGCGAGGCACGAGAGTGTCCACAGTTCCGCG 465
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      1886  AGTAGTAGTAGTGACTCGCGAGGCGGGATGTGCGGGACCGGCTGGCCCGAGGAGCCTGCT 1945
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Qy      466  G 466
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db      1946  G 1946
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

Search completed: January 10, 2003, 06:43:11
Job time : 98.8611 secs

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:14:32 ; Search time 2221.92 Seconds  
(without alignments)  
4781.561 Million cell updates/sec

Title: US-09-581-500b-12  
Perfect score: 656  
Sequence: 1 gccacaacacaaatgaaat.....ccacacgagaccccgaga 656

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pin:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	384.4	58.6	636	14 BQ807824	BQ807824 NISC_Kk10
2	142.8	21.8	930	14 BQ435307	BQ435307 AGENCOURT
3	137.4	20.9	270	9 AA904435	AA904435 ok07q11.s
4	126.4	19.3	244	9 AA889105	AA889105 am38h04.s
5	107	16.3	601	9 AL708543	AL708543 DKFZp686B
6	91.8	14.0	707	13 B1819086	B1819086 603033444

7	76.4	11.6	742	13	B1756040
8	70	10.7	896	12	BQ499418
9	67	10.2	925	14	BQ29418
c 10	63.2	9.6	925	17	CNS0091P
11	61.4	9.4	839	17	CNS004NB
c 12	61.4	9.4	1071	17	CNS000EM
c 13	56.2	8.6	844	17	CNS0052P
c 14	54.8	8.4	348	9	AL664001
15	54.6	8.3	691	3	B1547708
16	54.4	8.3	300	9	AA853497
c 17	54.4	8.3	456	9	AL554951
18	54	8.2	876	17	CNS03C63
19	53.8	8.2	1052	13	BM553599
c 20	53.8	8.2	1336	14	BM810024
c 21	53.6	8.2	1101	17	CNS0153F
c 22	53.2	8.1	764	17	AG060189
c 23	53.2	8.1	1101	17	CNS00KXY
24	53.2	8.1	1145	10	B8422200
25	53	8.1	548	9	AL597104
26	53	8.1	897	17	AG060530
c 27	53	8.1	912	17	CNS006N3
c 28	52.8	8.0	609	10	AV704516
c 29	52.8	8.0	773	13	B1892255
c 30	52.8	8.0	977	9	AL578905
c 31	52.8	8.0	1199	14	BQ706827
32	52.6	8.0	613	17	AG043036
c 33	52.6	8.0	750	17	CNS04PWH
34	52.2	8.0	677	17	AG154064
35	52.2	8.0	949	17	AG171092
36	52	7.9	925	12	BQ742444
c 37	51.8	7.9	430	12	BG655144
c 38	51.6	7.9	846	17	CNS00KOV
39	51.6	7.9	949	14	BQ940345
c 40	51.6	7.9	967	17	CNS01LMA
41	51.6	7.9	1036	17	CNS010BS
42	51.4	7.8	533	9	AL572608
c 43	51.4	7.8	798	13	B1951298
44	51.4	7.8	834	17	AG131468
45	51.2	7.8	450	9	AL513867

## ALIGNMENTS

RESULT 1  
BQ807824/c 636 bp mRNA linear EST 31-JUL-2002  
LOCUS NISC\_Kk10f04.y1 NCI\_CGAP\_Brn72 Macaca mulatta cDNA clone  
DEFINITION IMAGE:5331199.5', mRNA sequence.  
ACCESSION BQ807824  
VERSION BQ807824.1 GI:22032033  
KEYWORDS EST.  
SOURCE rhesus monkey.  
ORGANISM Macaca mulatta  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae;  
Cercopitheciinae; Macaca.  
REFERENCE 1 (bases 1 to 636)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NIHC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov  
Plate: LLAM11840 row: K column: 8  
Seq primer: M13Rpl reverse primer (ABI).

FEATURES  
source

```

Location/Qualifiers
1. .636
/organism="Macaca mulatta"
/db_xref="taxon:9544"
/clone="IMAGE:5331199"
/clone_lib="NCI_CGAP_Brn72"
/tissue_type="hypothalamus"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: Brain; Vector: pCMV-SPORT6, ccdB; Site: 1;
NotI; Site: 2; EcoRV; Cloned unidirectionally. Primer:
oligo dt. Average insert size 2.2 kb. Constructed by
in vitro. Note: this is a NCI_CGAP Library."

BASE COUNT      81 a      210 c      190 g      155 t
ORIGIN
Query Match      58.6%; Score 384.4; DB 14; Length 636;
Best Local Similarity 86.9%; Pred. No. 9.1e-85;
Matches 444; Conservative 1; Mismatches 60; Indels 6; Gaps 2;

QY 145 GAAAAAGGACAAATAGCAATGTGTACACTTTTATATAAAACCCCTCCCAAGGACGAG 204
      ||||| || || || || || || || || || || || || || || || || || || ||
Db 564 GAAAAACAAACCAATCTCTGTGTGACAAAGCCTTCACACACCCCTCCCAAGGACGAGT 505
      ||||| || || || || || || || || || || || || || || || || || || ||

QY 205 CACTGGCCCTCTCCGTTGCCACAGACATCCACAGGCCCCAAGAAATCAGGGATGTC 264
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 504 CACTGGCCCTCTCTCCGTTGCCACAGACATCCACAGGCCCCAAGAAATCAGGGATGTC 445
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 265 ACAAGCCAGAGCAATCAACAGGTTCTGAGTCATCTGCCGGAAGCCTTGCCCTCAATCAAG 324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 ACAGCCACAGCGCAATCGACAGGTTCTGAGTCATCTGCCGGAAGCCTTGCCCTCAATCAAG 385
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 325 CGGACGTGAAGCATCTACAAAGAGGAGTAATAGTCAAGACAGAGCGGCGGCGGCGGC 384
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 CGGTCGTGAAGCATCTACAAAGAGGAGTAATAGTCAAGACAGAGCGGCGGCGGCGGC 325
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 385 GCGACAGCAGCAGCAGCAGGAGGTGGGGGCTCTGCGCAGGTACCGGGGGGGGAGGCAC 444
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 GCGGCGCAGCAGCAGCAGGAG-----GTGGGGGGGCTCTGCCAGGTACCGGGGGGGAGGCAT 270
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 445 GGAGTGCCAGGTTCGCGGAGGCCACCTTCCTCCCTGGAGTGGTGAGAGAG--GGGAA 503
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 CGAGTGCCAGGTTCGCGGAGGCCACCTTCCTCCCTGGAGTGGTGAGAGAGTGGGGA 210
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 504 GGGAGGAGCCAGCAGCAGCAATCAGAGCGAGGCAAGAGCGGCGGCGGCACTANGAGATG 563
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 GAGAGGAGCCAAAGCAGCACTCAGAGCGAGGCAAGAGCGGAGGAGTAGGGGAATG 150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 564 ACSGCGGAGCGCGCGGGAAGAAATCTCGGGCTGTGGGGTCCCTGGCCACCGAC 623
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 ACGGTGGAGGCGCGCGGGAAGAGAGTCGCGGGGCTGTGGGGTCCCTGGCCACCGAC 90
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 624 CGGGTCCCAAGCCACCGCGAGACCCCGC 654
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 CGGGTCCCAAGCCACCGCGAGACCCCGC 59
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
BQ435307
LOCUS      BQ435307      930 bp      mRNA      linear      EST 24-MAY-2002
DEFINITION AGNCOURT_7926752 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6160773
5', mRNA sequence.
ACCESSION BQ435307
VERSION    BQ435307.1 GI:21174383
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.

```

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DFP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13511 row: p column: 22

High quality sequence stop: 607.

Location/Qualifiers

1. .930

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6160773"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/Note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI;

Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2 kb. Library constructed by Life

Technologies."

BASE COUNT 268 a 210 c 263 g 186 t 3 others

## ORIGIN

Query Match 21.8%; Score 142.8; DB 14; Length 930;

Best Local Similarity 98.6%; Pred. No. 6.1e-25;

Matches 144; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 262 TGCACAGCCAGAGCAATCGAAGCGTTCTGAGTCATCTGCCGGAAGCCTTGCCCTCAATC 321

Db 20 TCCGCAAGCCAGAGCAATCGAAGCGTTCTGAGTCATCTGCCGGAAGCCTTGCCCTCAATC 79

QY 322 AAGCGGACGTGAGCATCTACAAAGAGGAGTAATAGTCAAGACAGAGCGGCGGCGGC 381

Db 80 AAGCGGACGTGAGCATCTACAAAGAGGAGTAATAGTCAAGACAGAGCGGCGGCGGC 139

QY 382 GCGGCGCAGCAGCAGCAGCAGGAG 407

Db 140 GCGGCGCAGCAGCAGCAGGAG 165

RESULT 3

## LOCUS

AA904435

DEFINITION

ok07g11.s1 Soares\_NFL\_T-GBC\_S1 Homo sapiens cDNA clone

IMAGE:1507172 3', mRNA sequence.

ACCESSION

AA904435

VERSION

AA904435.1 GI:3039558

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 270)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 437 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 232.

Location/Qualifiers

1. .270

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1507172"

/clone\_lib="Soares\_NFL\_T-GBC\_S1"

/lab\_host="DH10B"

[illegible]

	QY	222	GTCGCCACATCATCACAG - GCCCAAGAATCAGGGATTGCACAGCCAGAGCAATC	280
	Dd	64	GTCGCCGAGACATCCACACAGGCCAAAGAATCAGGGATTGCACAAGCCAGAGCAATC	123
	QY	281	GAACGGTTCTGAGTCAATCTG	300
	Dd	124	GAACGGTTCTGAGTCAATCTG	143
 RESULT 5 AL708543				
	LOCUS	AL708543	601 bp	mRNA linear EST 22-MAR-2002
	DEFINITION	DKFZP86B1453_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone		
		DKFZP86B1453 5', mRNA sequence.		
	ACCESSION	AL708543		
	VERSION	AL708543.1	GI:19691898	
	KEYWORDS	EST.		

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Theria; Primates; Carnivora; Canidae; Canis; Canis lupus  
 1 (bases 1 to 601)  
 Wambutt, R., Heubner, D., Mewes, W., Weil, B. and Wiemann, S.  
 EST (Wambutt, R., Heubner, D., Mewes, W., Weil, B. and Wiemann, S.)  
 Unpublished (1999)  
 Contact: Wambutt R  
 MIPS  
 Am Klopferspitz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
 sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 No sl sequence available.  
 This clone (DKFZp68B1453) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```

source
FEATURES
Location/Qualifiers
1. .601
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686B1453"
/clone_lib="686 (synonym: hloc3)"

```

```
/tissue_type=human skeletal muscle  
/dev_stage="adult"  
/lab_host="DH0B"  
/notes="Vector: pTriplex2; Site_1: SfIIA; site_2: SfiIB;  
cdna-collection"
```

	BASE COUNT	167 a	133 c	167 g	134 t
ORIGIN					
Query Match					
Best Local Similarity					
Matches 107;					
Conservative					
0;					
Mismatches					
Indels					
Gaps					
Oy	318	AATCAAGCGGACCTGAAGCATC	TACAAGAAGGAGNATGAAAGCACGCAGCGGGCGG	377	
Ddb	1	AATCAAGCGGACCTGAAGCATC	TACAAGAAGGAGNATGAAAGCACGCAGCGGGCGG	60	
OY	378	CGGCGGCCGACGACGACGACGACGACGAGGTGGGGCCCTTCTGCCAG	424		

```

|||||
Db 61 CGCGGGCAGCAGCAGCAGCAGAGTGGGGCCTCTGCCAG 107

RESULT 6
BI819086 707 bp mRNA linear EST 04-OCT-2001
LOCUS 603033444F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5174458 5',
DEFINITION mRNA sequence.
ACCESSION BI819086
VERSION BI819086.1 GI:15930636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 707)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1434 row: h column: 11
High quality sequence stop: 697.
Location/Qualifiers
1. .707
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5174458"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 Kb,
insert size range 1-3 Kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."
BASE COUNT 208 a 149 c 191 g 159 t
ORIGIN

Query Match 14.0%; Score 91.8; DB 13; Length 707;
Best Local Similarity 97.9%; Pred. No. 2.6e-12;
Matches 93; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 330 CQTGAGCATCTCAAGAGGAGTAAGTCAAGCAGCAGCGCGGGCGGGCGGCAG 389
Db 1 CQTGAGCATCTCAAGAGGAGTAAGTCAAGCAGCAGCGCGGGCGGGCGGCAG 60

QY 390 CAGCAGCAGCAGCAGAGTGGGGCCTCTGCCAG 424
Db 61 CAGCAGCAGCAGCAGAGTGGGGCCTCTGCCAG 95

RESULT 7
BI756040 742 bp mRNA linear EST 25-SEP-2001
LOCUS 603030386F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200677 5',
DEFINITION mRNA sequence.
ACCESSION BI756040
VERSION BI756040.1 GI:15747618
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 742)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1502 row: 1 column: 22
High quality sequence stop: 742.
Location/Qualifiers
1. .742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5200677"
/clone_lib="NIH_MGC_114"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site.1: NotI;
Site.2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 Kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 223 a 153 c 200 g 166 t
ORIGIN

Query Match 11.6%; Score 76.4; DB 13; Length 742;
Best Local Similarity 98.7%; Pred. No. 1.7e-08;
Matches 77; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 330 CQTGAGCATCTCAAGAGGAGTAAGTCAAGCAGCAGCGCGGGCGGGCGGCAG 389
Db 1 CQTGAGCATCTCAAGAGGAGTAAGTCAAGCAGCAGCGCGGGCGGGCGGCAG 60

QY 390 CAGCAGCAGCAGCAGGAG 407
Db 61 CAGCAGCAGCAGCAGGAG 78

RESULT 8
BG499418 896 bp mRNA linear EST 27-MAR-2001
LOCUS 602546648F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4669282 5',
DEFINITION mRNA sequence.
ACCESSION BG499418
VERSION BG499418.1 GI:13460947
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be

```









Search completed: January 10, 2003, 06:15:15  
Job time : 2229.02 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:57:20 ; Search time 83.435 Seconds  
(without alignments)  
7673.783 Million cell updates/sec

Title: US-09-581-500B-13

Sequence: 22  
1 atcgaacgggttcgtgcatct 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg\_hum.\*
- 31: em.htg\_inv.\*
- 32: em.htg\_other.\*
- 33: em.htg\_mus.\*
- 34: em.htg\_pln.\*
- 35: em.htg\_rod.\*
- 36: em.htg\_mam.\*
- 37: em.htg\_vrt.\*
- 38: em.sy.\*
- 39: em.htgo\_hum.\*
- 40: em.htgo\_mus.\*
- 41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22	100.0	22	6	AX021007	AX021007 Sequence
2	22	100.0	656	6	AX021006	AX021006 Sequence
3	22	100.0	191395	9	AC009802	AC009802 Homo sapi
4	22	100.0	191793	2	AP001897	AP001897 Homo sapi
5	22	100.0	198291	2	AC067875	AC067875 Homo sapi
6	22	100.0	201734	2	AC090224	AC090224 Homo sapi
7	20.4	92.7	186351	2	AP001503	AP001503 Homo sapi
8	17.8	80.9	95448	9	AC005572	AC005572 Homo sapi
9	17.8	80.9	124092	2	CNS08C9K	AL732645 Oryza sat
10	17.8	80.9	175238	10	AL606903	AL606903 Mouse DNA
11	17.8	80.9	176641	2	AC084869	AC084869 Homo sapi
12	17.8	80.9	178103	2	AC118856	AC118856 Rattus no
13	17.8	80.9	228664	2	AL606971	AL606971 Mus muscu
14	17.8	80.9	258929	9	AC010433	AC010433 Homo sapi
15	17.4	79.1	192082	10	AL591490	AL591490 Mouse DNA
16	17.2	78.2	1163	8	AY088527	AY088527 Arabidops
17	17.2	78.2	1175	8	AFRNASCD	X90454 A.thaliana
18	17.2	78.2	1689	6	A75661	A75661 Sequence 6
19	17.2	78.2	2294	8	AF105034	AF105034 Arabidops
20	17.2	78.2	3004	8	AF069468	AF069468 Arabidops
21	17.2	78.2	3084	6	A75674	A75674 Sequence 19
22	17.2	78.2	3569	3	HC024941	AJ24941 haemochu
23	17.2	78.2	10174	9	HS598F21A	AL96857 Novel hum
24	17.2	78.2	13202	6	AX346386	AX346386 Sequence
25	17.2	78.2	13202	6	AX346387	AX346387 Sequence
26	17.2	78.2	99886	9	HS598F2	AL021579 Human DNA
27	17.2	78.2	103904	8	ATAC021640	AC021640 Arabidops
28	17.2	78.2	119670	2	AP003988	AP003988 Oryza sat
29	17.2	78.2	121622	2	AC124971	AC124971 Medicago
30	17.2	78.2	134282	2	AP004229	AP004229 Oryza sat
31	17.2	78.2	156302	2	AC123298	AC123298 Rattus no
32	16.8	76.4	1356	1	AF005745	AF005745 Unculture
33	16.8	76.4	55784	2	AC120157	AC120157 Mus muscu
34	16.8	76.4	58376	2	AC127705	AC127705 Rattus no
35	16.8	76.4	71290	2	AC123313	AC123313 Rattus no
36	16.8	76.4	187392	10	AL683890	AL683890 Mouse DNA
37	16.8	76.4	207358	2	AC094155	AL732494 Mus muscu
38	16.8	76.4	214225	2	AC094155	AC094155 Rattus no
39	16.4	74.5	41084	3	AC006603	AC006603 Caenorhab
40	16.4	74.5	46855	8	AC078898	AC078898 Arabidops
41	16.4	74.5	98890	2	AC130896	AC130896 Homo sapi
42	16.4	74.5	166458	2	AC084737	AC084737 Homo sapi
43	16.4	74.5	178012	2	AC131214	AC131214 Homo sapi
44	16.4	74.5	191150	2	AC129402	AC129402 Rattus no
45	16.4	74.5	263876	9	AC019233	AC019233 Homo sapi

ALIGNMENTS

RESULT 1	AX021007	AX021007	22 bp	DNA	linear	PAT 07-SEP-2000
AX021007	Sequence	Sequence	13 from Patent WO9932643.			
LOCUS	AX021007	AX021007				
DEFINITION	AX021007	AX021007				
ACCESSION	AX021007.1	GI:10044670				
VERSION						
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
	Enkaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 22)					
AUTHORS	Del-Favero,J., Raeymaekers,P. and Van Broeckhoven,C.					
TITLE	Mood disorder gene					
JOURNAL	Patent: WO 9932643-A 13 01-JUL-1999;					

BROECKHOVEN CHRISTINE VAN (BE); DEL FAVERO JURGEN (BE); RAEMYMAEKERS  
PETER (BE); VLAAMS INTERUNIV INST BIOTECH (BE)

# FEATURES

source  
1. 22  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 5 a 5 c 5 g 7 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best local Similarity 100.0%; Pred. No. 0.12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22

Db 1 ATCGAACGGTCTGAGTCATCT 22

## RESULT 2

LOCUS AX021006 656 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 12 from Patent WO9932643.  
ACCESSION AX021006  
VERSION AX021006.1 GI:10044669  
KEYWORDS  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 656)  
AUTHORS Del-Favero,J., Raemymaekers,P. and Van Broeckhoven,C.  
TITLE Mood disorder gene  
JOURNAL Patent: WO 9932643-A 12 01-JUL-1999;  
BROECKHOVEN CHRISTINE VAN (BE); DEL FAVERO JURGEN (BE); RAEMYMAEKERS  
PETER (BE); VLAAMS INTERUNIV INST BIOTECH (BE)

# FEATURES

source  
1. 656  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 183 a 165 c 208 g 96 t 4 others  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 656;  
Best local Similarity 100.0%; Pred. No. 0.18;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22

Db 278 ATCGAACGGTCTGAGTCATCT 299

## RESULT 3

LOCUS AC009802/c 191395 bp DNA linear PRI 17-FEB-2000  
DEFINITION Homo sapiens chromosome 18, clone RP11-793J2, complete sequence.  
ACCESSION AC009802  
VERSION AC009802.13 GI:6705901  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 191395)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Homo sapiens chromosome 18, clone RP11-793J2  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191395)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,

Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meldrum,J., Molla,M., Morrow,J., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

## Direct Submission

Submitted (01-SEP-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 191395)

# REFERENCE

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,  
Bouhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepe,X., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Miranda,C., Mlungu,V., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,  
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,  
Riley,K., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.

## Direct Submission

Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 16, 2000 this sequence version replaced gi:6514043.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: LI250

Center clone name: 793\_J\_2

# FEATURES

Location/Qualifiers

1. 191395

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-793J2"

/clone\_lib="RPCI-11 Human Male BAC"

2. 542

/rpt\_family="L1M2"

repeat\_region 474..708

repeat\_region /rpt\_family="L1"

repeat\_region 708..1353

repeat\_region /rpt\_family="L1"

repeat\_region 1351..1623

repeat\_region /rpt\_family="L1M4A"

repeat\_region 1632..1687

unsure /rpt\_family="(TATATG)n"

complement(1673..1726)

/note="Single-stranded coverage."

repeat\_region 1688..2207

repeat\_region /rpt\_family="L1M4A"

unsure complement(1724..1770)

```

repeat_region /note="Single-stranded terminator coverage."
repeat_region 2208..2215 /rpt_family="(TATATG)n"
repeat_region 2594..2648 /rpt_family="LIME"
repeat_region complement(2649..2991) /rpt_family="MER2"
repeat_region 2992..3475 /rpt_family="LIME"
repeat_region complement(4330..4437) /rpt_family="FRAM"
repeat_region 4438..4567 /rpt_family="(TA)n"
repeat_region 4599..4959 /rpt_family="AT_rich"
repeat_region 4982..5151 /rpt_family="(TATG)n"
repeat_region 5144..5311 /rpt_family="(TATAA)n"
repeat_region 5257..5441 /rpt_family="(TA)n"
repeat_region 5444..5525 /rpt_family="AT_rich"
repeat_region 5533..5708 /rpt_family="(TTATA)n"
repeat_region 5742..5907 /rpt_family="(TTATA)n"
repeat_region complement(5966..6047) /rpt_family="FRAM/FRAM"
repeat_region 6682..6703 /rpt_family="AT_rich"
repeat_region 7034..7933 /rpt_family="L1P16"
repeat_region 7935..8006 /rpt_family="AluY"
repeat_region 8009..9111 /rpt_family="L1P16"
repeat_region 9112..9165 /rpt_family="AT_rich"
repeat_region 9216..9496 /rpt_family="AluSx"
repeat_region 9497..9520 /rpt_family="(CAAA)n"
repeat_region 9961..10264 /rpt_family="AluY"
repeat_region complement(10837..10840) /note="Single-stranded terminator coverage."
repeat_region complement(11156..11515) /rpt_family="THEIC"
repeat_region 11856..11987 /rpt_family="MIR"
repeat_region 12592..12643 /rpt_family="L2"
repeat_region 13683..13994 /rpt_family="AluSx"
repeat_region complement(13996..14160) /rpt_family="L1MEC"
repeat_region 14788..15255 /rpt_family="MER67b"
repeat_region complement(15875..15996) /rpt_family="MER5B"
repeat_region 16101..16135 /rpt_family="AT_rich"
repeat_region 16136..16304 /rpt_family="L1M4"
repeat_region complement(16572..16647) /rpt_family="MADE1"
repeat_region complement(16716..17011) /rpt_family="AluY"
repeat_region 17045..17780 /rpt_family="L1ME3"
repeat_region 17827..17938 /rpt_family="L1ME3"

```

```

repeat_region complement(17945..18133) /rpt_family="MER58A"
repeat_region 18222..18755 /rpt_family="LIME3"
repeat_region 18843..19019 /rpt_family="LIME3"
repeat_region 19573..19616 /rpt_family="(CACG)n"
repeat_region 20506..20544 /rpt_family="AT_rich"
repeat_region 21894..22184 /rpt_family="AluSq"
repeat_region complement(22787..22998) /rpt_family="LFR40a"
repeat_region complement(23074..23133) /rpt_family="LFR40b"
repeat_region 23184..23540 /rpt_family="L2"
repeat_region complement(23675..25225) /rpt_family="pTR5"
repeat_region 25229..25582 /rpt_family="THELB"
repeat_region 25879..26031 /rpt_family="MLT1A1"
repeat_region 26032..26321 /rpt_family="AluSx"
repeat_region 26322..26531 /rpt_family="MLT1A1"
repeat_region 26623..26736 /rpt_family="L2"
repeat_region 26943..27439 /rpt_family="L2"
repeat_region 27914..27985 /rpt_family="MIR"

```

Query Match 100.0%; Score 22; DB 9; Length 191395;

Best Local Similarity 100.0%; Pred. No. 0.37;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCGACGGTCTGAGTCATCT 22

Db 97997 ATCGACGGTCTGAGTCATCT 97976

RESULT 4

AP001897 191793 bp DNA linear HTG 30-MAY-2000  
 LOCUS Homo sapiens chromosome 18 clone RP11-693A18 map 18q22, WORKING  
 DEFINITION DRAFT SEQUENCE, 54 unordered pieces.

ACCESSION AP001897.2 GI:8117548  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Homo sapiens DNA, clone:RP11-693A18.  
 SOURCE ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 191793)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Homo sapiens 191,793 genomic DNA of 18q22

JOURNAL Published Only in Database (2000)

REFERENCE 2 (bases 1 to 191793)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,

Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE Direct Submission

JOURNAL Submitted (24-Apr-2000) Masahira Hattori, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,

Japan (E-mail:hattori@gsc.riken.go.jp,

URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,

Fax:81-42-778-9924)

COMMENT On May 30, 2000 this sequence version replaced gi:7649784.

----- Genome Center

Center: RIKEN Genomic Sciences Center(GSC)

Center code: RIKEN

Web site: <http://ngp.gsc.riken.go.jp/>

Contact: [hattori@ngs.riken.go.jp](mailto:hattori@ngs.riken.go.jp)

----- Project Information

Center project name: HumDraft18

Center clone name: Rp11-693A18

----- Summary Statistics

Sequencing vector: PCR products; 100% of reads

Chemistry: Dye-terminator ET-amersham; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 167461 bases at least Q40

Consensus quality: 174712 bases at least Q30

Consensus quality: 180769 bases at least Q20

Insert size: 186493; sum-of-contigs

Quality coverage: 4.60x in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently consists of 54 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1  
10824 10723 contig of 10723 bp in length  
19551 19450 contig of 8627 bp in length  
28840 28739 contig of 9189 bp in length  
37897 37796 contig of 8957 bp in length  
45744 45643 contig of 7477 bp in length  
53301 53200 contig of 7457 bp in length  
59567 59466 contig of 6166 bp in length  
65124 65023 contig of 5457 bp in length  
71726 71725 contig of 6602 bp in length  
77046 77045 contig of 5221 bp in length  
82059 82058 contig of 4913 bp in length  
86135 86134 contig of 4356 bp in length  
90937 90936 contig of 4221 bp in length  
95358 95357 contig of 4321 bp in length  
99377 99376 contig of 4375 bp in length  
103852 103851 contig of 4556 bp in length  
109508 109507 contig of 3961 bp in length  
112569 112568 contig of 3843 bp in length  
116512 116511 contig of 4179 bp in length  
120791 120790 contig of 4999 bp in length  
125890 125889 contig of 3540 bp in length  
129530 129529 contig of 3346 bp in length  
132976 132975 contig of 3743 bp in length  
136819 136818 contig of 3386 bp in length  
140305 140304 contig of 2908 bp in length  
143313 143312 contig of 2999 bp in length  
146412 146411 contig of 2283 bp in length  
148795 148794 contig of 2757 bp in length  
151652 151651 contig of 1820 bp in length  
153572 153571 contig of 2334 bp in length  
155905 155904 contig of 1925 bp in length  
156006 156005 contig of 2305 bp in length  
158031 158030 contig of 1522 bp in length  
160436 160435 contig of 2058 bp in length  
162058 162057 contig of 2159 bp in length  
164215 164214 contig of 2160 bp in length  
166473 166472 contig of 2062 bp in length  
168735 168734 contig of 1835 bp in length  
170897 170896 contig of 1633 bp in length  
172832 172831 contig of 1116 bp in length  
174565 174564 contig of 1138 bp in length  
175781 175780 contig of 1804 bp in length  
177019 177018 contig of 1155 bp in length  
178923 178922 contig of 1071 bp in length  
180178 180177 contig of 1151 bp in length  
181349 181348 contig of 207 bp in length  
182600 182606 contig of 1086 bp in length  
182907 182907 contig of 1106 bp in length  
184093 184093 contig of 1106 bp in length

185299 186320 contig of 1022 bp in length  
186421 187601 contig of 1181 bp in length  
187702 189099 contig of 1398 bp in length  
189200 190576 contig of 1377 bp in length  
190677 191793 contig of 1117 bp in length  
Sequence updated (26-May-2000).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 54 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
10723 10723: contig of 10723 bp in length  
10824 10823: gap of 100 bp  
10824 19450: contig of 8627 bp in length  
19451 19550: gap of 100 bp  
28739 28739: contig of 9189 bp in length  
28740 28839: gap of 100 bp  
37796 37796: contig of 8957 bp in length  
37897 37896: gap of 100 bp  
45644 45743: gap of 100 bp  
53200 53300: contig of 7457 bp in length  
53301 59466: contig of 6166 bp in length  
59467 59566: gap of 100 bp  
59567 65023: contig of 5457 bp in length  
65024 65123: gap of 100 bp  
71725 71725: contig of 6602 bp in length  
71726 71825: gap of 100 bp  
77046 77046: contig of 5221 bp in length  
77147 82059: contig of 4913 bp in length  
82060 82159: gap of 100 bp  
82160 86515: contig of 4356 bp in length  
86516 86615: gap of 100 bp  
86616 90836: contig of 4221 bp in length  
90837 90936: gap of 100 bp  
90937 95257: contig of 4321 bp in length  
95258 95357: gap of 100 bp  
95358 99276: contig of 3919 bp in length  
99277 99376: gap of 100 bp  
99377 103751: contig of 4375 bp in length  
103752 103851: gap of 100 bp  
103852 108407: contig of 4556 bp in length  
108408 108507: gap of 100 bp  
108508 112468: contig of 3961 bp in length  
112469 112568: gap of 100 bp  
112569 116411: contig of 3843 bp in length  
116412 116511: gap of 100 bp  
116512 120690: contig of 4179 bp in length  
120691 120790: gap of 100 bp  
120791 125789: contig of 4999 bp in length  
125790 125889: gap of 100 bp  
125890 129429: contig of 3540 bp in length  
129430 129529: gap of 100 bp  
129530 132875: contig of 3346 bp in length  
132876 132975: gap of 100 bp  
132976 136718: contig of 3743 bp in length  
136719 136818: gap of 100 bp  
136819 140204: contig of 3386 bp in length  
140205 140304: gap of 100 bp  
140305 143212: contig of 2908 bp in length  
143213 143312: gap of 100 bp  
143313 146311: contig of 2999 bp in length  
146312 146411: gap of 100 bp  
146412 148694: contig of 2283 bp in length  
148695 148794: gap of 100 bp  
148795 151551: contig of 2757 bp in length  
151552 151651: gap of 100 bp  
151652 153471: contig of 1820 bp in length

Barna, N., Bastien, V., Boguslavskiy, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Choepe, Y., Colangelo, M., Collins, S.,

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslawski, L., Boukhgalter, B., Brown, A., Burkett, G.,  
Campobiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collimore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, N., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaCocque, K., Lamazares, R., Landers, T., Lehoczyk, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Ollivar, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollars, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Rothmann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Tridillo, J.,

misc_feature	1. 45183	/note="assembly_fragment clone_end:SP6 vector_side:left"	/map="l8"
misc_feature	45284. 46768	/note="assembly_fragment"	/clone="RP11-693A18"
misc_feature	46869. 47961	/note="assembly_fragment"	/clone_lib="RPCI-11 Human Male BAC"
misc_feature	48062. 49643	/note="assembly_fragment"	
misc_feature	49744. 51393	/note="assembly_fragment"	
misc_feature	51494. 74531	/note="assembly_fragment"	
misc_feature	74632. 127710	/note="assembly_fragment"	
misc_feature	127811. 201734	/note="assembly_fragment"	
		clone_end:P7	
BASE COUNT	61611	vector_side:right"	708
ORIGIN	a 39444 c 39572 g 60399 t		

```
Query Match 100.0%; Score 22; DB 2; Length 201734;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 ATCGAACGGTCTCTGAGTCACT 22  
DB 147792 ATCGAACGGTCTCTGAGTCACT 147771

Db 147792 ATCGAACGGTCTGAGTCATCT 147771

## RESULT 7

REC'D  
AP001503

LOCUS AP001503 186351 bp D

**DEFINITION** Homo sapiens chromosome 18 clone RP1

DRAFT SEQUENCE, 23 unordered pieces.

ACCESSION AP001503

VERSION AP001503.2 GI:8117361

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens DNA, clone:RP11-879N20.

ORGANISM      Homo sapiens

Eukaryota; Metazoa;  
Mammalia; Eutheria;  
Eukaryota; Chordata; Cranial

1 (bases 1 to 186351)  
Mammalia; Eutheria; Primates; Catari

REFERENCE	AUTHORS	Y	T	I	(bases 1 to 186351)	T	Y
	Hattori M						
	Tsibiki K						
	Tovoda N						

**AUTHORS**  
HALLGÖLL, M., ISHII, K., TOYODA, A., TAY  
ENRIJIMA, A., YADA, T., TOTOKI, Y., WAT

**TITLE** Homo sapiens 186,351 genomic DNA of Fujiyama, A.; Iada, I.; Ioloxi, I.; Wat

JOURNAL  
of  
herpetology  
100/351 genomic DNA of  
published only in DataBase (2000)

REFERENCE  
2 (bases 1 to 186351)  
2 (bases 1 to 186351)

**AUTHORS**  
Hattori, M., Ishii, K., Toyoda, A., Tay

Fujiyama, A., Yada, T., Totoki, Y., Wat

**TITLE** Direct Submission

JOURNAL Submitted (17-MAR-2000) Masahira Hat

and Chemical Research (RIKEN), Genom

Kitasato Univ., 1-15-1 Kitasato, Sag

Japan (E-mail: hattori@gsc.riken.go.jp)

URL: <http://hgp.gsc.riken.go.jp/>, Tel  
Fax: 81-43-778-0024

Fax: 81-42-778-9924)

COMMENT: on May 30, 2000 this sequence version  
----- Genoma Center

----- Genome Center  
Center: PTKEN Genetic Sciences C

Center: RIKEN Genomic Sciences C  
Center code: RIKEN

Center code: RIKEN  
Web site: <http://hmr.nci.nih.gov>

Web site: <http://nyp.gsc.tken.org>  
Contact: [battori@gsc.tken.org](mailto:battori@gsc.tken.org)

```

contact: haco@ikeyse.ikeyn.gu.jp
----- project Information

```

Center project name: HumDraft18

Center project name: RANBIACT0  
Center clone name: RP11-879N20

----- Summary Statistics

Sequencing vector: PCR products;

1

Chemistry: Dye-terminator ET-amersham; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 162162 bases at least Q40  
 Consensus quality: 175389 bases at least Q30  
 Consensus quality: 181326 bases at least Q20  
 Insert size: 184151; sum-of-contigs  
 Quality coverage: 4.54x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 42519 contig of 42519 bp in length
42620 59852 contig of 17233 bp in length
59953 74496 contig of 14544 bp in length
74597 87440 contig of 12844 bp in length
87541 99577 contig of 10382 bp in length
99678 110659 contig of 10982 bp in length
110760 115413 contig of 4654 bp in length
115514 124990 contig of 9477 bp in length
125091 135744 contig of 10654 bp in length
135845 145972 contig of 10128 bp in length
146073 153170 contig of 7098 bp in length
153271 163644 contig of 3128 bp in length
163745 173030 contig of 5226 bp in length
173031 181636 contig of 3864 bp in length
181637 183644 contig of 3910 bp in length
183645 184761 contig of 3756 bp in length
184762 186351 contig of 2110 bp in length
186352 177303 contig of 1439 bp in length
177304 179097 contig of 1694 bp in length
179098 181636 contig of 2439 bp in length
181637 183644 contig of 2439 bp in length
183645 184761 contig of 1908 bp in length
184762 186351 contig of 1017 bp in length
186352 177303 contig of 1490 bp in length

```

Sequence updated (26-May-2000).

\* NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 42519: contig of 42519 bp in length
42620 42619: gap of 100 bp
42620 59852: contig of 17233 bp in length
59853 59952: gap of 100 bp
59953 74496: contig of 14544 bp in length
74497 74596: gap of 100 bp
74597 87440: contig of 12844 bp in length
87441 87440: gap of 100 bp
87441 99577: contig of 12037 bp in length
99578 99677: gap of 100 bp
99678 110659: contig of 10982 bp in length
110660 110759: gap of 100 bp
110760 115413: contig of 4654 bp in length
115414 115513: gap of 100 bp
115514 124990: contig of 9477 bp in length
124991 125090: gap of 100 bp
125091 135744: contig of 10654 bp in length
135745 135844: gap of 100 bp
135845 145972: contig of 10128 bp in length
145973 146072: gap of 100 bp
146073 153170: contig of 7098 bp in length
153171 153270: gap of 100 bp
153271 156398: contig of 3128 bp in length
156399 156498: gap of 100 bp
156499 161724: contig of 5226 bp in length
161725 161824: gap of 100 bp

```

```

* 161825 165688: contig of 3864 bp in length
* 165689 165788: gap of 100 bp
* 165789 169698: contig of 3910 bp in length
* 169699 169798: gap of 100 bp
* 169799 173554: contig of 3756 bp in length
* 173555 173654: gap of 100 bp
* 173655 175764: contig of 2110 bp in length
* 175765 175864: gap of 100 bp
* 175865 177303: contig of 1439 bp in length
* 177304 177403: gap of 100 bp
* 177404 179097: contig of 1694 bp in length
* 179098 179197: gap of 100 bp
* 179198 181636: contig of 2439 bp in length
* 181637 181736: gap of 100 bp
* 181737 183644: contig of 1908 bp in length
* 183645 183744: gap of 100 bp
* 183745 184761: contig of 1017 bp in length
* 184762 184861: gap of 100 bp
* 184862 186351: contig of 1490 bp in length.

```

# FEATURES

```

Source
1. .186351
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="18"
   /map="18q22"
   /clone="RP11-879N20"
1. .42519
   /note="assembly_fragment"
42620. 59852
   /note="assembly_fragment"
59953. 74496
   /note="assembly_fragment"
74597. 87440
   /note="assembly_fragment"
87541. 99577
   /note="assembly_fragment"
99678. 110659
   /note="assembly_fragment"
110760. 115413
   /note="assembly_fragment clone_end:SP6 vector_side:right"
115514. 124990
   /note="assembly_fragment"
125091. 135744
   /note="assembly_fragment"
135845. 145972
   /note="assembly_fragment"
146073. 153170
   /note="assembly_fragment"
153271. 156398
   /note="assembly_fragment clone_end:T7 vector_side:left"
156499. 161724
   /note="assembly_fragment"
161825. 165688
   /note="assembly_fragment"
165789. 169698
   /note="assembly_fragment"
169799. 173554
   /note="assembly_fragment"
173655. 175764
   /note="assembly_fragment"
175865. 177303
   /note="assembly_fragment"
177404. 179097
   /note="assembly_fragment"
179198. 181636
   /note="assembly_fragment"
181737. 183644
   /note="assembly_fragment"
183745. 184761
   /note="assembly_fragment"
184862. 186351
   /note="assembly_fragment"

```

BASE COUNT 56377 a 36459 c 35952 g 55363 t 2200 others



STS

```

/repeat_region /standard_name="4let7"
/db_xref="dbSTS:G31407"
45099..45386
/rpt_family="Alu"
/misc_feature complement(48265..48384)
/note="GRAIL 2 excellent exon, frame 0"
/misc_feature complement(4958..50004)
/note="GRAIL 2 excellent exon, frame 0"
/repeat_region complement(50482..50554)
/rpt_family="Alu"
/repeat_region complement(50903..50972)
/rpt_family="Alu"
/repeat_region complement(51301..51570)
/rpt_family="Alu"
/repeat_region 53075..53337
/rpt_family="Alu"
/misc_feature complement(59112..59259)
/note="GRAIL 2 excellent exon, frame 0"
/repeat_region 61720..61983
/rpt_family="Alu"
/repeat_region 61998..62033
/note="(AATA)9"
/rpt_type=tandem
/misc_feature 66046..66147
/rpt_unit=AATA
/repeat_region 68813..68852
/note="(ATCT)10"
/rpt_type=tandem
/misc_feature complement(68888..68992)
/rpt_family="MIR2"
/misc_feature 72177..72322
/standard_name="CDC1e3"
/note="99% identity B07638 (exon trapped product)"
/repeat_region 72614..72649
/note="(AC)18"
/rpt_type=tandem
/repeat_region complement(73385..73695)
/rpt_family="Alu"
/misc_feature complement(73801..73925)
/note="8% identity EST oe52h01.s1"
/db_xref="dbEST:AA862321"
73944..94270
/standard_name="LTR7"
/repeat_region 76779..77132
/rpt_family="THE1"
/repeat_region 77616..77670
/rpt_family="MADE1"
/repeat_region complement(80901..81095)
/rpt_family="Alu"
/misc_feature 81076..81185
/note="GRAIL 2 excellent exon, frame 2"
/repeat_region complement(81404..81485)
/rpt_family="MIR"
/repeat_region complement(82147..82531)
/rpt_family="THE1"
/misc_feature complement(82908..83342)
/note="99% identity EST Yg82c08.r1"
/db_xref="dbEST:R53285"
83752..84040
/rpt_family="Alu"
/repeat_region 84023..84044
/note="(A)22"
/rpt_type=tandem

```

Query Match 80.9% Score 17.8; DB 9; Length 95448;  
 Best Local Similarity 90.5%; Pred. NO. 99;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0;

QY 2 TCAGACGGTCTGAGTCATCT 22

```

Db 44786 TCAGACGGTCTGAGTCATCT 44766
|||||
CNS08C9K 124092 bp DNA linear HTG 04-JUN-2002
Oryza sativa chromosome 12 clone OSUNBB0090H23, *** SEQUENCING IN
PROGRESS ***, 11 ordered pieces.
ACCESSION AL732645
VERSION AL732645.1 GI:21326737
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 124092)
AUTHORS Cholsne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
Segurens,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
Weissenbach,J. and Quetier,F.
Oryza sativa chromosome 12 sequencing
Unpublished
Genoscope.
Direct Submission
Submitted (03-JUN-2002) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence.
Work on the sequence is in progress and the release of this data is
based on the understanding that the sequence may change as work
continue. The sequence may be contaminated with foreign sequence
from E.coli, yeast, vector, phage, etc.
Contigs composition :
1707 bp contig from 1 to 1707
13309 bp contig from 1808 to 15116
6358 bp contig from 15217 to 21574
22606 bp contig from 21675 to 44280
9040 bp contig from 44381 to 53420
20348 bp contig from 53521 to 73868
9188 bp contig from 73969 to 83156
19261 bp contig from 83257 to 102517
6512 bp contig from 102618 to 109129
8922 bp contig from 109230 to 118151
5841 bp contig from 118252 to 124092.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequences as soon as it is available and
* the accession number will be preserved.
* 1 1707: contig of 1707 bp in length
* 1708 1807: gap of 100 bp
* 1808 15116: contig of 13309 bp in length
* 15117 15216: gap of 100 bp
* 15217 21574: contig of 6358 bp in length
* 21575 21674: gap of 100 bp
* 21675 44280: contig of 22606 bp in length
* 44281 44380: gap of 100 bp
* 44381 53420: contig of 9040 bp in length
* 53421 53520: gap of 100 bp
* 53521 73868: contig of 20348 bp in length
* 73869 73968: gap of 100 bp
* 73969 83156: contig of 9188 bp in length
* 83157 83256: gap of 100 bp
* 83257 102517: contig of 19261 bp in length
* 102518 102617: gap of 100 bp
* 102618 109129: contig of 5512 bp in length

```

\* 109130 109229: gap of 100 bp  
 \* 109230 118151: contig of 8922 bp in length  
 \* 118152 118251: gap of 100 bp  
 \* 118252 124092: contig of 5841 bp in length.

## FEATURES

Location/Qualifiers  
 1..124092  
 /organism="Oryza sativa"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:4530"  
 /chromosome="12"  
 /clone="OSJNB0090H23"  
 /clone.lib="OSJNBb"

BASE COUNT 33707 a 27560 c 27075 g 34750 t 1000 others  
 ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 124092;  
 Best Local Similarity 90.5%; Pred. No. 1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTGAGTCATC 21  
 |||||

Db 90490 ACCGACGGTCTGAGTCATC 90510

## RESULT 10

AL606903 175238 bp DNA linear ROD 20-FEB-2002  
 LOCUS Mouse DNA sequence from clone Rp23-27813 on chromosome 4, complete  
 DEFINITION  
 AL606903  
 VERSION AL606903.8 GI:18873508  
 KEYWORDS HTG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

AUTHORS Garner,P.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humberv@sanger.ac.uk Clone requests: clonesrequest@sanger.ac.uk

## COMMENT

On Feb 22, 2002 this sequence version replaced gi:17381401.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em:, EMBL; Sw:,  
 SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP  
 database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep  
 from the RCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACe3.6  
 This sequence is the entire insert of clone Rp23-27813.

FEATURES  
 source

1..175238  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="4"  
 /clone="Rp23-27813"

BASE COUNT 45250 a 39075 c 40628 g 50285 t  
 ORIGIN

Query Match 80.9%; Score 17.8; DB 10; Length 175238;  
 Best Local Similarity 90.5%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCGACGGTCTGAGTCATCT 22  
 |||||

Db 11296 TCGATCGGTTCTGAGCATCT 11316

## RESULT 11

AC084869/c 176641 bp DNA linear HTG 11-MAR-2001  
 LOCUS Homo sapiens chromosome RFI-11 clone RP11-508E4, WORKING DRAFT  
 DEFINITION  
 AC084869  
 VERSION AC084869.2 GI:11465172  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 176641)  
 AUTHORS Waterston,R.H.  
 TITLE The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 176641)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-NOV-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## COMMENT

On Nov 30, 2000 this sequence version replaced gi:11323446.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----  
 Center project name: H.NH0508E04

----- Summary Statistics -----  
 Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator Big Dye; 0% of reads

Chemistry: Dye-primer ET; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 142830 bases at least Q40

Consensus quality: 154079 bases at least Q30

Consensus quality: 160559 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 172341; sum-of-contigs

Quality coverage: 2.73 in Q20 bases; agarose-fp

Quality coverage: 3.02 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 44 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1667: contig of 1667 bp in length  
 \* 1668 1767: gap of unknown length  
 \* 1768 3037: contig of 1270 bp in length  
 \* 3038 3137: gap of unknown length  
 \* 3138 4370: contig of 1233 bp in length  
 \* 4371 4471: gap of unknown length  
 \* 4471 6093: contig of 1622 bp in length  
 \* 6093 6192: gap of unknown length

\* 6193 7578: contig of 1386 bp in length  
 \* 7579 7678: gap of unknown length  
 \* 9287: contig of 1609 bp in length  
 \* 9288 9387: gap of unknown length  
 \* 11041: contig of 1654 bp in length  
 \* 11042 11141: gap of unknown length  
 \* 11142 12827: contig of 1686 bp in length  
 \* 12828 12927: gap of unknown length  
 \* 12928 15022: contig of 2095 bp in length  
 \* 15023 15122: gap of unknown length  
 \* 15123 16154: contig of 1032 bp in length  
 \* 16155 16255: gap of unknown length  
 \* 16255 17925: contig of 1670 bp in length  
 \* 17925 18025: gap of unknown length  
 \* 18025 20170: contig of 2146 bp in length  
 \* 20171 20270: gap of unknown length  
 \* 20271 21811: contig of 2911 bp in length  
 \* 21812 22281: gap of unknown length  
 \* 22282 23440: contig of 2059 bp in length  
 \* 23441 25441: gap of unknown length  
 \* 25441 28590: contig of 3150 bp in length  
 \* 28591 28690: gap of unknown length  
 \* 28691 31015: contig of 2325 bp in length  
 \* 31016 31115: gap of unknown length  
 \* 31116 33667: contig of 2752 bp in length  
 \* 33668 37010: contig of 3043 bp in length  
 \* 37011 39680: contig of 2570 bp in length  
 \* 39681 39781: gap of unknown length  
 \* 39782 42432: contig of 2652 bp in length  
 \* 42433 42533: gap of unknown length  
 \* 42533 45045: contig of 2512 bp in length  
 \* 45045 45144: gap of unknown length  
 \* 45145 48026: contig of 2882 bp in length  
 \* 48027 48126: gap of unknown length  
 \* 48127 51134: contig of 3008 bp in length  
 \* 51135 51235: gap of unknown length  
 \* 51235 53679: contig of 2445 bp in length  
 \* 53680 53779: gap of unknown length  
 \* 53780 56546: contig of 2767 bp in length  
 \* 56547 56646: gap of unknown length  
 \* 56647 59359: contig of 2713 bp in length  
 \* 59360 59459: gap of unknown length  
 \* 59460 62921: contig of 3462 bp in length  
 \* 62922 63021: gap of unknown length  
 \* 63022 65458: contig of 2437 bp in length  
 \* 65459 65559: gap of unknown length  
 \* 65560 68534: contig of 3976 bp in length  
 \* 68535 69635: gap of unknown length  
 \* 69635 73068: contig of 3434 bp in length  
 \* 73069 73169: gap of unknown length  
 \* 73170 78010: contig of 4842 bp in length  
 \* 78011 83482: contig of 5372 bp in length  
 \* 83483 83582: gap of unknown length  
 \* 83583 88800: contig of 5218 bp in length  
 \* 88801 89900: gap of unknown length  
 \* 89901 97207: contig of 8307 bp in length  
 \* 97208 97308: gap of unknown length  
 \* 97309 104462: contig of 7155 bp in length  
 \* 104463 104562: gap of unknown length  
 \* 104563 111426: contig of 6864 bp in length  
 \* 111427 111526: gap of unknown length  
 \* 111527 118832: contig of 7306 bp in length  
 \* 118833 118932: gap of unknown length  
 \* 118933 125694: contig of 6762 bp in length  
 \* 125695 125794: gap of unknown length  
 \* 125795 133108: contig of 7314 bp in length  
 \* 133109 133208: gap of unknown length  
 \* 133209 141040: contig of 7832 bp in length  
 \* 141041 141140: gap of unknown length  
 \* 141141 150417: contig of 9277 bp in length

\* 150418 150517: gap of unknown length  
 \* 150518 158426: contig of 7909 bp in length  
 \* 158427 158526: gap of unknown length  
 \* 158527 167566: contig of 9040 bp in length  
 \* 167567 167666: gap of unknown length  
 \* 167667 176641: contig of 8975 bp in length.  
 FEATURES  
 source  
 1..176641  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="RP11-11"  
 /clone="RP11-508E4"  
 1..1667  
 /note="assembly\_name:Contig15"  
 1768..3037  
 /note="assembly\_name:Contig16"  
 3138..4370  
 /note="assembly\_name:Contig18"  
 4471..6092  
 /note="assembly\_name:Contig19"  
 6193..7578  
 /note="assembly\_name:Contig20"  
 7679..9287  
 /note="assembly\_name:Contig21"  
 9388..11041  
 /note="assembly\_name:Contig22"  
 11142..12827  
 /note="assembly\_name:Contig23"  
 12928..15022  
 /note="assembly\_name:Contig24"  
 15123..16154  
 /note="assembly\_name:Contig25"  
 16255..17924  
 /note="assembly\_name:Contig26"  
 18025..20170  
 /note="assembly\_name:Contig27"  
 20271..23181  
 /note="assembly\_name:Contig28"  
 23282..25340  
 /note="assembly\_name:Contig29"  
 25441..28590  
 /note="assembly\_name:Contig30"  
 28691..31015  
 /note="assembly\_name:Contig31"  
 31116..33667  
 /note="assembly\_name:Contig32"  
 33968..37010  
 /note="assembly\_name:Contig33"  
 37111..39680  
 /note="assembly\_name:Contig34"  
 39781..42432  
 /note="assembly\_name:Contig35"  
 42533..45044  
 /note="assembly\_name:Contig36"  
 45145..48026  
 /note="assembly\_name:Contig37"  
 48127..51134  
 /note="assembly\_name:Contig38"  
 51235..53679  
 /note="assembly\_name:Contig39"  
 53780..56546  
 /note="assembly\_name:Contig40"  
 56647..59359  
 /note="assembly\_name:Contig41"  
 59460..62921  
 /note="assembly\_name:Contig42"  
 63022..65458  
 /note="assembly\_name:Contig43"  
 65559..69534

Query Match 80.9%; Score 17.8; DB 2; Length 176641;  
 Best Local Similarity 90.5%; Pred.No.1.1e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TCGAACGGTTCTGAGTCATCT 22  
||| ||| ||| ||| ||| ||| |||  
Db 76238 TCTAAGGGTTCTGAGTCATCT 76218

RESULT 12					
LOCUS	AC118856/c				
DEFINITION	Rattus norvegicus clone CH230-246K21, *** SEQUENCING IN PROGRESS				
	***, 47 unordered pieces.				
ACCESSION	AC118856				
VERSION	AC118856.2				
KEYWORDS	GI:21747306				
SOURCE	HTG_PHASE1.				
ORGANISM	Norway rat.				
	Rattus norvegicus				
	Eukaryota; Metazoa;				
	Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia;				
	Sciurognathi; Muridae; Murinae;				
REFERENCE	Rattus.				
	1 (bases 1 to 178103)				

```

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20258320.  
 ----- Genome Center -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 ----- Project Information -----

```

Center project name: GWMO
Center clone name: CH230-246X21
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 137677 bases at least Q40
Consensus quality: 144197 bases at least Q30
Consensus quality: 148410 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*

```

44171 44270: gap of unknown length  
44271 46570: contig of 2300 bp in length  
46571 46670: gap of unknown length  
46671 49009: contig of 3239 bp in length  
49009 50009: gap of unknown length  
50009 53419: contig of 3410 bp in length  
53419 53519: gap of unknown length  
53519 53820: contig of 2868 bp in length  
53820 56487: gap of unknown length  
56487 60933: contig of 4445 bp in length  
60933 61033: gap of unknown length  
61033 63607: contig of 2575 bp in length  
63607 63707: gap of unknown length  
63707 69387: contig of 5680 bp in length  
69387 69487: gap of unknown length  
69487 71963: contig of 2476 bp in length  
71963 72063: gap of unknown length  
72063 75957: contig of 3894 bp in length  
75957 76057: gap of unknown length  
76057 79569: contig of 3512 bp in length  
79569 83877: contig of 6208 bp in length  
83877 85977: gap of unknown length  
85977 90718: contig of 4741 bp in length  
90718 90818: gap of unknown length  
90818 94719: contig of 3900 bp in length  
94719 94818: gap of unknown length  
94818 100271: contig of 5453 bp in length  
100271 100372: gap of unknown length  
100372 106250: contig of 5879 bp in length  
106250 106350: gap of unknown length  
106350 112115: contig of 5765 bp in length  
112115 112215: gap of unknown length  
112215 119503: contig of 7288 bp in length  
119503 119603: gap of unknown length  
119603 125419: contig of 5816 bp in length  
125419 125519: gap of unknown length  
125519 133159: contig of 7640 bp in length  
133159 133259: gap of unknown length  
133259 140941: contig of 7682 bp in length  
140941 141041: gap of unknown length  
141041 147717: contig of 6676 bp in length  
147717 147817: gap of unknown length  
147817 157811: contig of 9994 bp in length  
157811 157911: gap of unknown length  
157911 169168: contig of 11257 bp in length  
169168 169268: gap of unknown length  
169268 178103: contig of 8835 bp in length.  
Location/Qualifiers  
1. .178103  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-246K21"

BASE COUNT 47068 a 39502 c 39309 g 47550 t 4674 others  
ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 178103;  
Best Local Similarity 90.5%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGAGTCATC 21  
||||| ||||| |||||  
Db 161889 ATCGAACGGTTCGAGTCATC 161869

RESULT 13  
AL606971 228664 bp DNA linear HTG 13-AUG-2002  
LOCUS  
DEFINITION Mus musculus chromosome 4 clone RP23-123120, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 2 unordered pieces.  
ACCESSION AL606971  
VERSION AL606971.9 GI:21689948  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.

SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 228664)  
AUTHORS Ramsay H.  
JOURNAL Direct Submission  
TITLE Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 3, 2002 this sequence version replaced gi:21261813.  
COMMENT ----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BM123120  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752: 100% of reads  
Chemistry: Dye-terminator; 2% of reads  
Chemistry: Dye-terminator Big Dye; 97% of reads  
Consensus quality: 228120 bases at least Q40  
Consensus quality: 228283 bases at least Q30  
Consensus quality: 228383 bases at least Q20  
Insert size: 228564; sum-of-contigs  
Insert size: 203349; 5.9% error; agarose-fp  
Quality coverage: 11.87x in Q20 bases; sum-of-contigs Quality  
coverage: 13.78x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 138728: contig of 138728 bp in length  
\* 138729 138828: gap of 100 bp  
\* 138829 228664: contig of 89836 bp in length.  
Location/Qualifiers  
1. .228664  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-123120"  
/clone\_lib="RPCI-23"  
misc\_feature 1. .138728  
/note="assembly\_fragment:02064"  
misc\_feature 138829. .228664  
/note="assembly\_fragment:07697"  
BASE COUNT 56606 a 56012 c 56424 g 59522 t 100 others  
ORIGIN

FEATURES  
source

misc\_feature  
1. .138728  
/note="assembly\_fragment:02064"

misc\_feature  
138829. .228664  
/note="assembly\_fragment:07697"

Query Match 80.9%; Score 17.8; DB 2; Length 228664;  
Best Local Similarity 90.5%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATCT 22  
||||| ||||| ||||| |||||  
Db 51570 TCGAACGGTTCGAGTCATCT 51590

RESULT 14  
AC010433/c 258929 bp DNA linear PRI 10-JUL-2002  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-2202L20, complete sequence.  
ACCESSION AC010433  
VERSION AC010433.9 GI:21717106  
KEYWORDS HTG.  
SOURCE human.



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 19:55:35 ; Search time 10.4381 Seconds  
(without alignments)  
4746.432 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atogaacggttgtgagtcattc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq 101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	22	100.0	22	20	AA38557	Human chromosome 1
2	22	100.0	656	20	AA38553	Human chromosome 1
3	17.2	78.2	486	20	AAV87361	EST clone BV275.
4	17.2	78.2	1164	21	AA41897	Arabidopsis thalia
5	17.2	78.2	1522	21	AAC77253	Human ORFX ORF2808
6	17.2	78.2	1689	14	AAQ52498	Helminth aminopept
7	17.2	78.2	3006	14	AAQ52491	Helminth aminopept
8	17.2	78.2	3084	14	AAQ52500	Helminth aminopept
9	17.2	78.2	10347	22	AAF28060	Human 07CG27 gene

10	17.2	78.2	13202	24	ABL33484	Human immune syste
11	17.2	78.2	13202	24	ABL33485	Human immune syste
12	16.2	73.6	388	21	AAC27333	Human secreted pro
13	16.2	73.6	458	21	AA40395	Ze mays DNA fragm
14	16.2	73.6	470	23	ABV13083	Human prostate exp
15	16.2	73.6	911	21	AA43639	Ze mays DNA fragm
16	16.2	73.6	1030	21	AAC34348	Arabidopsis thalia
17	16.2	73.6	1394	9	AA46621	Synthetic Hydroxym
18	16.2	73.6	1394	14	AA46621	Encodes human 3-hy
19	16.2	73.6	15320	23	ABL07834	Drosophila melanog
20	15.8	71.8	1203	24	ABN68044	Streptococcus poly
21	15.8	71.8	1293	24	ABK72739	Bacillus lichenifo
22	15.8	71.8	1527	21	AAC43261	Arabidopsis thalia
23	15.8	71.8	1662	24	ABL41878	Nucleotide sequenc
24	15.6	70.9	292	24	ABL71087	Corn tassal-derive
25	15.6	70.9	398	21	AAA79317	Eucalyptus grandis
26	15.6	70.9	461	21	AAA79313	Eucalyptus grandis
27	15.6	70.9	1266	21	AAC51603	Arabidopsis thalia
28	15.6	70.9	1499	21	AAC39132	Arabidopsis thalia
29	15.6	70.9	1619	17	AA118796	Human immunophilin
30	15.6	70.9	2447	15	AAQ56702	Partial sequence o
31	15.6	70.9	2447	22	AA111038	Human kappa opioid
32	15.6	70.9	2719	23	ABL28558	Drosophila melanog
33	15.6	70.9	3109	21	AAC78150	Human cancer assoc
34	15.6	70.9	3528	22	AAS23393	Candida albicans e
35	15.6	70.9	147419	24	ABK83574	Human cDNA differe
36	15.6	70.9	325791	22	AAS43104	Human Oestrogen re
37	15.6	70.9	1830121	17	AA42063	Haemophilus influe
38	15.6	70.9	2944528	24	ABA03041	Listeria monocytog
39	15.4	70.0	723	23	AAS66765	DNA encoding novel
40	15.4	70.0	1117	21	AAC35713	Arabidopsis thalia
41	15.4	70.0	1117	21	AAC51417	Arabidopsis thalia
42	15.4	70.0	1650	22	AA159082	Human polynucleoti
43	15.4	70.0	1711	19	AAV33136	Plasmodium berghei
44	15.4	70.0	1735	22	AAK52200	Human polynucleoti
45	15.4	70.0	2168	21	AAC59479	Human secreted pro

#### ALIGNMENTS

#### RESULT 1

AA38557

ID AAX38557 standard; DNA; 22 BP.

XX AAX88557;

AC AAX88557;

XX AAX88557;

DT 10-SEP-1999 (first entry)

XX Human chromosome 18q YAC clone amplification primer.

DE Human chromosome 18q; mood disorder; polymorphic marker; detection;

XX identification; trinucleotide repeat expansion; schizophrenia;

KW anxiety disorder; adjustment disorder; personality disorder;

XX nucleotide triplet repeat; ss.

XX Synthetic.

OS Homo sapiens.

XX WO9932643-A2.

XX 01-JUL-1999.

PD 17-DEC-1998; 98WO-EP08543.

XX 18-DEC-1997; 97GB-0026804.

PR (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PA Del-Favero J, Raeymaekers P, Van Broeckhoven C;

XX WPI; 1999-418934/35.

XX



PT Detecting nucleotide triplet repeats in human chromosome 18q

PS Claim 29; Fig 15b; 87pp; English.

XX The present invention describes detecting nucleotide triplet repeats in  
 CC a region of human chromosome 18q disposed between polymorphic markers  
 CC D18S68 and D18S979 to identify a human gene associated with a mood  
 CC disorder or related disorder. AAX88542 to AAX88705 represents human  
 CC chromosome 18q YAC clones and primers corresponding to them, used in the  
 CC exemplification of the present invention. YAC clones comprising a  
 CC portion of the region of human chromosome 18q between markers D18S68 and  
 CC D18S979 are used to identify at least one human gene associated with a  
 CC mood disorder or related disorder. The mood disorder or related  
 CC disorder, is chosen from the Diagnostic and Statistical Manual of Mental  
 CC Disorders, version 4 (DSM-IV) taxonomy. This includes mood disorders  
 CC (296.XX, 300.4, 311, 301, 13, 295.70), schizophrenia and related  
 CC disorders (295, 297.1, 298.9, 297.3, 298.9), anxiety disorders (300.XX,  
 CC 309.81, 308.3), adjustment disorders (309.XX) and personality disorders  
 CC (codes 301.XX). Probes derived from genes associated with the mood  
 CC disorder or related disorder can be used to detect pathological  
 CC mutations or genetic variations in patients. The methods, probes and  
 CC antibodies can be used to determine the susceptibility of an individual  
 CC to a mood disorder or related disorder. The nucleic acids and proteins  
 CC of the human gene can be used to treat mood disorders and related  
 CC disorders.

XX Sequence 22 BP; 5 A; 5 C; 5 G; 7 T; 0 other;

Query Match 100.0%; Score 22; DB 20; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 0.074;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTCGAGTCATCT 22

DB 1 ATCGACGGTCTCGAGTCATCT 22

RESULT 2

AAX88553

ID AAX88553 standard; DNA; 656 BP.

XX AC AAX88553;

XX 10-SEP-1999 (first entry)

XX Human chromosome 18q YAC clone nucleotide sequence #12.

XX Human chromosome 18q; mood disorder; polymorphic marker; detection;  
 KW identification; trinucleotide repeat expansion; schizophrenia;  
 KW anxiety disorder; adjustment disorder; personality disorder;  
 KW nucleotide triplet repeat; ss.

XX Homo sapiens.

XX Synthetic.

XX WO9932643-A2.

XX 01-JUL-1999.

XX 17-DEC-1998; 98WO-BP08543.

XX 18-DEC-1997; 97GB-0026804.

XX (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

XX Del-Favero J, Raeymaekers P, Van Broeckhoven C;

XX WPI; 1999-418934/35.

XX Detecting nucleotide triplet repeats in human chromosome 18q

XX Claim 21; Fig 15a; 87pp; English.

CC The present invention describes detecting nucleotide triplet repeats in  
 CC a region of human chromosome 18q disposed between polymorphic markers  
 CC D18S68 and D18S979 to identify a human gene associated with a mood  
 CC disorder or related disorder. AAX88542 to AAX88705 represents human  
 CC chromosome 18q YAC clones and primers corresponding to them, used in the  
 CC exemplification of the present invention. YAC clones comprising a  
 CC portion of the region of human chromosome 18q between markers D18S68 and  
 CC D18S979 are used to identify at least one human gene associated with a  
 CC mood disorder or related disorder. The mood disorder or related  
 CC disorder, is chosen from the Diagnostic and Statistical Manual of Mental  
 CC Disorders, version 4 (DSM-IV) taxonomy. This includes mood disorders  
 CC (296.XX, 300.4, 311, 301, 13, 295.70), schizophrenia and related  
 CC disorders (295, 297.1, 298.9, 297.3, 298.9), anxiety disorders (300.XX,  
 CC 309.81, 308.3), adjustment disorders (309.XX) and personality disorders  
 CC (codes 301.XX). Probes derived from genes associated with the mood  
 CC disorder or related disorder can be used to detect pathological  
 CC mutations or genetic variations in patients. The methods, probes and  
 CC antibodies can be used to determine the susceptibility of an individual  
 CC to a mood disorder or related disorder. The nucleic acids and proteins  
 CC of the human gene can be used to treat mood disorders and related  
 CC disorders.

XX Sequence 656 BP; 183 A; 165 C; 208 G; 96 T; 4 other;

Query Match 100.0%; Score 22; DB 20; Length 656;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTCTCGAGTCATCT 22

DB 278 ATCGACGGTCTCGAGTCATCT 299

RESULT 3

AAV87361

ID AAV87361 standard; cDNA; 486 BP.

XX AC AAV87361;

XX 27-APR-1999 (first entry)

XX EST clone BV275.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US06954.

XX 10-APR-1997; 97US-0835913.

XX (GEMY) GENETICS INST INC.

XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;

XX Racie IA, Spaulding V, Treacy M;

XX WPI; 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries

XX Claim 1; Page 540; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are

CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.

XX SQ Sequence 486 BP; 145 A; 114 C; 110 G; 116 T; 1 other;

Query Match 78.28; Score 17.2; DB 20; Length 486;  
 Best Local Similarity 86.48; Pred. No. 35;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATCGAAGGTTCTGAGTCATCT 22  
 | ||||| ||||| ||||| ||  
 Db 18 ACCGAACAGTCTGAGTCATTT 39

## RESULT 4

RAC41897  
 ID AAC41897 standard; DNA; 1164 BP.

XX AC RAC41897;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 33540.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

KW metabolic pathway; promoter; terminator; termination sequence; ss.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000BP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139452.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 23-JUN-1999; 99US-0140354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0140823.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144003.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 22-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 Query Match 78.2%; Score 17.2; DB 21; Length 1164;  
 Best Local Similarity 86.4%; Pred. No. 39;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 ATCGAACGGTTCGATCATCT 22  
 Db 201 ACCGAATCGTTCGATCATCT 222  
 RESULT 5  
 AAC77253  
 ID AAC77253 standard; cDNA; 1522 BP.  
 AC AAC77253;  
 XX  
 XX 08-FEB-2001 (first entry)  
 DT Human ORFX ORF2808 polynucleotide sequence SEQ ID NO:5615.  
 DE Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;  
 KW vulnery; antipapillary; antipapillary; antipapillary; antipapillary;  
 KW anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;  
 KW thrombosis; contraceptive; ss.  
 XX Homo sapiens.  
 OS  
 XX WC2000058473-A2.  
 XX  
 XX 05-OCT-2000.  
 PD  
 XX 31-MAR-2000; 2000WO-US08621.  
 XX  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach M;  
 PI  
 XX WPI; 2000-602362/57.  
 DR P-PSDB; AAB43044.  
 DR  
 XX Novel nucleic acids and peptides derived from open reading frame X,  
 PT useful for treating e.g. cancers, proliferative disorders,  
 PT neurodegenerative disorders and cardiovascular disease -  
 XX  
 PS Claim 5; Page 4799-4800; 5507pp; English.  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipapillary; antipapillary; antipapillary; antipapillary;  
 CC osteopathic; anticonvulsant; antipapillary; antipapillary; antipapillary;  
 CC

xx The sequences given in AAQ52498-99 represent cDNA clones of the helminth  
 CC aminopeptidase gene H11-1. H11-1 encodes H110D which is a protein  
 CC doublet which shows homology to a family of integral membrane amino-  
 CC peptidases. The differences between these coding sequences can be  
 CC attributed to different mRNAs of the multigene family, and also to  
 CC different variants of the H110D-encoding sequence being present at  
 CC different stages of the life cycle, or in strains differing in

CC The sequences given in AAQ52489-91 represent the helminth amino-  
CC peptidase genes HLI-3, -2 and -1 respectively. These sequences  
CC encode HLI00 which is a protein doublet which shows homology to a  
CC family of integral membrane aminopeptidases. The differences  
CC between these coding sequences can be attributed to different mRNAs  
CC of the multigene family, and also to different variants of the HLI00  
CC encoding sequence being present at different stages of the life  
CC cycle, or in strains differing in geographical origin. Antigenic  
CC fragments of the aminopeptidases encoded by these sequences may be  
CC used in vaccines to stimulate immune response against helminth  
CC parasites in humans or other animals. These DNA sequences may be  
CC incorporated into a virus or microbe and used in a similar manner.

XX SQ Sequence 3006 BP; 932 A; 600 C; 716 G; 758 T; 0 other;  
 Query Match 78.2%; Score 17.2; DB 14; Length 3006;  
 Best Local Similarity 86.4%; Pred. No. 44;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTCTTGAGTCATCT 22  
 ||||| ||||| ||||| |||||  
 Db 1429 ATCGAAGCGTCTTGAGTCATCT 1408

RESULT 8  
 AAF2500/c  
 ID AAF2500 standard; DNA; 3084 BP.  
 XX AC AAF2500;  
 XX DT 31-MAY-1994 (first entry)  
 XX DE Helminth aminopeptidase H11-1 gene.  
 XX KW Helminth; aminopeptidase; H11-3; H11-2; H11-1; integral membrane;  
 KW alpha-amino acyl peptide hydrolase (microsome); multigene family;  
 KW antigen; vaccine; parasite; human; virus; microbe; ss.  
 XX OS Haemonchus contortus.  
 XX FH Key Location/Qualifiers  
 FT CDS 23..2956  
 FT /\*tag= a  
 FT /product= H110D variant  
 XX WO9323542-A.  
 XX PD 25-NOV-1993.  
 XX PF 07-MAY-1993; 93WO-GB00943.  
 XX PR 08-MAY-1992; 92GB-0009993.  
 XX PA (AGRI-) AGRIC & FOOD RES COUNCIL.  
 XX PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JU;  
 XX PT Smith TS;  
 XX WPI; 1993-386574/48.  
 XX DR P-PSDB; AAF51282.  
 XX PS New DNA encoding amino-peptidase from Haemonchus contortus and  
 FT derived peptide(s) - useful in protective vaccines, for new  
 FT vectors, transformed cells and oligosaccharide, for incorporation  
 FT in virus or microbe  
 XX Claim 1; Page 75; 137pp; English.

The sequences given in AAF2500-02 represent the helminth amino-peptidase genes H11-1, -2 and -3 respectively. These sequences encode H110D which is a protein doublet which shows homology to a family of integral membrane aminopeptidases. The differences between these coding sequences can be attributed to different mRNAs of the multigene family, and also to different variants of the H110D-encoding sequence being present at different stages of the life cycle, or in strains differing in geographical origin. Antigenic fragments of the aminopeptidases encoded by these sequences may be used in vaccines to stimulate immune response against helminth parasites in humans or other animals. These DNA sequences may be incorporated into a virus or microbe and used in a similar manner.

XX SQ Sequence 3084 BP; 955 A; 610 C; 729 G; 790 T; 0 other;  
 Query Match 78.2%; Score 17.2; DB 14; Length 3084;  
 Best Local Similarity 86.4%; Pred. No. 44;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTCTTGAGTCATCT 22  
 ||||| ||||| ||||| |||||  
 Db 1429 ATCGAAGCGTCTTGAGTCATCT 1408

RESULT 9  
 AAF28060  
 ID AAF28060 standard; CDNA; 10347 BP.  
 XX AC AAF28060;  
 XX DT 23-MAY-2001 (first entry)  
 XX DE Human 07CG27 gene cDNA.  
 XX KW Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer;  
 KW oncogene; ss.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 225..8684  
 FT /\*tag= a  
 FT /product= "07CG27 protein"  
 XX WO200116291-A2.  
 XX PD 08-MAR-2001.  
 XX PF 25-AUG-2000; 2000WO-US23291.  
 XX PR 27-AUG-1999; 99US-0151049.  
 XX PA (MYRI-) MYRIAD GENETICS INC.  
 PA (HOSP-) HOSPITAL FOR SICK CHILDREN.  
 XX PT Tavtigian SV, Swedlund B, Simard J, Rommens JM;  
 XX WPI; 2001-226682/23.  
 XX DR P-PSDB; AAB35408.  
 XX PT Novel human prostate cancer marker gene termed as 07CG27 gene, useful  
 PT for screening mutations in the gene in diagnosis of a predisposition to  
 PT cancer -  
 XX Claim 2; Page 79-91; 99pp; English.

The present invention provides the protein and coding sequences of the human 07CG27 oncogene. This gene is found at the HPC1 region of chromosome 1. The sequences can be used in the diagnosis and identification of treatments for prostate cancer. The present sequence is the 07CG27 coding sequence.

XX SQ Sequence 10347 BP; 3325 A; 2451 C; 2229 G; 2342 T; 0 other;  
 Query Match 78.2%; Score 17.2; DB 22; Length 10347;  
 Best Local Similarity 86.4%; Pred. No. 52;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTCTTGAGTCATCT 22  
 ||||| ||||| ||||| |||||  
 Db 2567 ACCGAACAGTCTTGAGTCATTT 2588

RESULT 10  
 ABL33484  
 ID ABL33484 standard; DNA; 13202 BP.  
 XX AC ABL33484;  
 XX DT 26-MAR-2002 (first entry)

```

XX DE Human immune system associated gene SEQ ID NO: 1457.
XX DE
XX KW Human: immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytotatic; nootropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
XX KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX KW gene; ds.
XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP07537.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX DR
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation -
XX PS Claim 1; SEQ ID NO 1457; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX SQ Sequence 13202 BP; 3470 A; 187 C; 2897 G; 6648 T; 0 other;
XX
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX SQ Sequence 13202 BP; 3470 A; 187 C; 2897 G; 6648 T; 0 other;
XX
Query Match 78.2%; Score 17.2; DB 24; Length 13202;
Best Local Similarity 86.4%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTTCTGAGTCACT 22
Db 4944 ATCGAAGCGTTCTGAGTCACT 4965

RESULT 11
ABL33485/c
ID ABL33485 standard; DNA; 13202 BP.
XX AC ABL33485;
XX
XX DT 26-MAR-2002 (first entry)
XX DE Human immune system associated gene SEQ ID NO: 1458.
XX KW Human: immune system disease; cytosine methylation; antiasthmatic;
XX KW antiarteriosclerotic; antianaemic; cytotatic; nootropic;
XX KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX KW antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
XX KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
XX KW gene; ds.

```

```

XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP07537.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-130909/17.
XX DR
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation -
XX PS Claim 1; SEQ ID NO 1458; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX SQ Sequence 13202 BP; 3984 A; 187 C; 2851 G; 6180 T; 0 other;
XX
Query Match 78.2%; Score 17.2; DB 24; Length 13202;
Best Local Similarity 86.4%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTTCTGAGTCACT 22
Db 8259 ATCGAAGCGTTCTGAGTCACT 8238

RESULT 12
AAC27333/c
ID AAC27333 standard; cDNA; 388 BP.
XX AC AAC27333;
XX
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein 5' EST, SEQ ID NO: 31408.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping; ss.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.
XX PR 21-FEB-2000; 2000EP-0200610.
XX PR 26-FEB-1999; 99US-0122487.
XX PA (GEST ) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX DR WPI; 2000-500381/45.
XX

```

```

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX
XX Claim 1; SEQ ID 31408; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or poly(A+) RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX SQ Sequence 388 BP; 130 A; 55 C; 80 G; 123 T; 0 other;
Query Match 73.6%; Score 16.2; DB 21; Length 388;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAAGGTCGAGTCATCT 22
   || || || || || || || || ||
Db 290 TCTAAGTCGAGTCATCT 270

RESULT 13
AAC40995
ID AAC40995 standard; DNA; 458 BP.
AC AAC40995;
XX
XX 17-OCT-2000 (first entry)
DE
DE Zea mays DNA fragment SEQ ID NO: 30271.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
XX Zea mays subsp. mays.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 21-MAY-1999; 99US-0135629.
XX 24-MAY-1999; 99US-0136021.
XX 25-MAY-1999; 99US-0136392.
XX 27-MAY-1999; 99US-0136782.
XX 28-MAY-1999; 99US-0137222.
XX 01-JUN-1999; 99US-0137528.
XX 03-JUN-1999; 99US-0137502.
XX 04-JUN-1999; 99US-0137724.
XX 07-JUN-1999; 99US-0138094.
XX 08-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 10-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139899.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 16-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145087.

```

PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 73.6%; Score 16.2; DB 21; Length 458;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0.

QY 2 TCGACGGCTCTGAGTCATCT 22  
||||| ||||| | |||||  
Db 362 TCGACGGCTCGGGTCATCT 382

RESULT 14  
ABV13083  
ID ABV13083 standard; cDNA; 470 BP.  
XX  
AC ABV13083;  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA I3074.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
XX Claim 1; Page 2160; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer



[illegible]

PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 73.6%; Score 16.2; DB 21; Length 911;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCTGAGTCATCT 22  
||||| ||||| ||||| ||||| |||||  
Db 370 TCGACCGGTTCCGGGTCATCT 390

Search completed: January 9, 2003, 21:33:51  
Job time : 16.4381 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:50:31 ; Search time 1.85235 Seconds  
(without alignments)  
3642.332 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgaaaggttctgagtcattc 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.2	78.2	1589	3	US-08-335-844A-6
C 2	17.2	78.2	3084	3	US-08-335-844A-19
C 3	15.6	70.9	398	4	US-09-228-986-55
C 4	15.6	70.9	461	4	US-09-228-986-51
C 5	15.6	70.9	2447	4	US-08-387-707-12
C 6	15.6	70.9	2447	4	US-08-405-271A-12
C 7	15.2	69.1	1436	4	US-09-270-542-91
C 8	15.2	69.1	1445	4	US-09-270-542-89
C 9	15.2	69.1	1709	4	US-09-270-542-87
C 10	15.2	69.1	2436	4	US-09-270-542-85
C 11	14.8	67.3	802	4	US-09-724-864-12
C 12	14.6	66.4	1248	4	US-09-134-001C-1980
C 13	14.6	66.4	1378	4	US-09-149-476-208
C 14	14.6	66.4	1473	2	US-08-541-033A-25
C 15	14.6	66.4	1473	2	US-08-828-451-25
C 16	14.6	66.4	1506	2	US-08-541-033A-23
C 17	14.6	66.4	1506	2	US-08-828-451-23
C 18	14.6	66.4	1740	4	US-08-991-944-1
C 19	14.6	66.4	1969	2	US-08-541-033A-7
C 20	14.6	66.4	1969	2	US-08-828-451-7
C 21	14.6	66.4	2096	2	US-08-541-033A-19
C 22	14.6	66.4	2096	2	US-08-828-451-19
C 23	14.6	66.4	2099	2	US-08-541-033A-3
C 24	14.6	66.4	2099	2	US-08-828-451-3
C 25	14.6	66.4	2137	2	US-08-541-033A-18
C 26	14.6	66.4	2137	2	US-08-828-451-18
C 27	14.6	66.4	2140	2	US-08-541-033A-1

Sequence 1, Appli  
Sequence 14, Appli  
Sequence 24, Appli  
Patent No. 5240838  
Patent No. 5240838  
Sequence 2, Appli  
Sequence 204, App  
Sequence 5, Appli  
Sequence 58, Appli  
Sequence 3, Appli  
Sequence 7, Appli  
Sequence 72, Appli  
Sequence 22, Appli  
Sequence 22, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 7, Appli  
Sequence 2, Appli

2140 2 US-08-828-451-1  
2319 4 US-09-134-078-14  
2323 4 US-09-149-476-24  
144 6 5240838-13  
173 6 5240838-12  
427 1 US-07-925-920-2  
1083 4 US-08-858-207A-204  
1505 4 US-09-193-377B-5  
2055 4 US-09-134-078-58  
2688 3 US-08-458-434A-3  
9299 3 US-08-458-434A-7  
46819 4 US-09-453-702B-72  
31 4 US-08-664-962B-22  
31 4 US-09-311-743-22  
845 4 US-09-453-323-4  
892 1 US-08-241-465B-7  
892 1 US-08-241-465B-7  
1006 1 US-08-241-465B-2

ALIGNMENTS

RESULT 1

US-08-335-844A-6/c

; Sequence 6, Application US/08335844A

; Patent No. 6066503

; GENERAL INFORMATION:

; APPLICANT: GRAHAM, MARGARET

; APPLICANT: SMITH, TEEVOR STANLEY

; APPLICANT: MUNN, EDWARD ALBERT

; APPLICANT: KNOX, DAVID PATRICK

; APPLICANT: OLIVER, JOANNA JANE

; APPLICANT: NEWTON, SUSAN ELIZABETH

; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING

; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF

; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS

; NUMBER OF SEQUENCES: 73

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rothwell, Figg, Ernst & Kurz

; STREET: Suite 701-E, 555 Thirteenth St., N.W

; CITY: Washington

; STATE: D. C.

; COUNTRY: U.S.A.

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/335,844A

; FILING DATE: 09-JAN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB PCT/GB93/00943

; FILING DATE: 06-MAY-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9209936

; FILING DATE: 08-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: WALKER, Barbara W.

; REGISTRATION NUMBER: 35,400

; REFERENCE/DOCKET NUMBER: 1181-223A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)783-6040

; TELEFAX: (202)783-6031

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1689 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

```

; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-335-844A-6

Query Match 78.28; Score 17.2; DB 3; Length 1689;
Best Local Similarity 86.48; Pred. No. 5.1;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
    ||||| ||||| |||||
Db 505 ATCGAACGGTCTGATACATCT 484

RESULT 2
US-08-335-844A-19/c
; Sequence 19, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MUNN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: Suite 701-E, 535 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,844A
; FILING DATE: 09-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9209936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3084 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-335-844A-19

Query Match 78.28; Score 17.2; DB 3; Length 3084;
Best Local Similarity 86.48; Pred. No. 5.6;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-09-228-986-55/c

Query Match 70.94; Score 15.6; DB 4; Length 398;
Best Local Similarity 81.88; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
    ||||| ||||| ||||| ||
Db 292 ATCGAACGGTCTGATCCACT 271

RESULT 4
US-09-228-986-51/c
; Sequence 51, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-228-986-55

Query Match 70.94; Score 15.6; DB 4; Length 398;
Best Local Similarity 81.88; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTGAGTCATCT 22
    ||||| ||||| ||||| ||
Db 292 ATCGAACGGTCTGATCCACT 271

RESULT 5
US-08-387-707-12/c
; Sequence 12, Application US/08387707
; Patent No. 6265563
; GENERAL INFORMATION:
; APPLICANT: EVANS, CHRISTOPHER J.
; APPLICANT: KEITH, DUANE E.
; TITLE OF INVENTION: OPIOID RECEPTOR GENES
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue, N.W. Suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
- APPLICATION NUMBER: US/08/387,707  
FILING DATE: 10-SEP-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20526.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-387-707-12

Query Match 70.9%; Score 15.6; DB 4; Length 2447;  
Best Local Similarity 81.8%; Pred. No. 37;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGTGACTCATCT 22  
||| | ||||| |||||  
Db 750 ATCAAGTGGTTCGTGACTCATCT 729

RESULT 6  
US-08-405-271A-12/c  
Sequence 12, Application US/08405271A  
Patent No. 6432652  
GENERAL INFORMATION:  
APPLICANT: EVANS, CHRISTOPHER J.  
APPLICANT: KEITH, DUANE E.  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 PENNSYLVANIA AVENUE, NW, Suite 5500  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,271A  
FILING DATE: 14-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 22000-20526.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030 MRSNFOERSWSH  
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-405-271A-12

Query Match 70.9%; Score 15.6; DB 4; Length 2447;  
Best Local Similarity 81.8%; Pred. No. 37;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGTGACTCATCT 22  
||| | ||||| |||||  
Db 750 ATCAAGTGGTTCGTGACTCATCT 729

RESULT 7  
US-09-270-542-91/c  
Sequence 91, Application US/09270542  
Patent No. 6322976  
GENERAL INFORMATION:  
APPLICANT: Aitman, Timothy  
APPLICANT: Scott, James  
TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
TITLE OF INVENTION: Therapy  
FILE REFERENCE: 4198/78179  
CURRENT APPLICATION NUMBER: US/09/270,542  
CURRENT FILING DATE: 1999-03-17  
EARLIER APPLICATION NUMBER: 09/221,222  
EARLIER FILING DATE: 1999-12-23  
NUMBER OF SEQ ID NOS: 207  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 91  
LENGTH: 1436  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
US-09-270-542-91

Query Match 69.1%; Score 15.2; DB 4; Length 1436;  
Best Local Similarity 85.0%; Pred. No. 56;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGTGACTCATC 21  
||| | ||||| |||||  
Db 1104 TCGATAGGTTCTGACATC 1085

RESULT 8  
US-09-270-542-89/c  
Sequence 89, Application US/09270542  
Patent No. 6322976  
GENERAL INFORMATION:  
APPLICANT: Aitman, Timothy  
APPLICANT: Scott, James  
TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
TITLE OF INVENTION: Therapy  
FILE REFERENCE: 4198/78179  
CURRENT APPLICATION NUMBER: US/09/270,542  
CURRENT FILING DATE: 1999-03-17  
EARLIER APPLICATION NUMBER: 09/221,222  
EARLIER FILING DATE: 1999-12-23  
NUMBER OF SEQ ID NOS: 207  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 89  
LENGTH: 1445  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
US-09-270-542-89  
Query Match 69.1%; Score 15.2; DB 4; Length 1445;  
Best Local Similarity 85.0%; Pred. No. 56;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATC 21  
 |||| |||||  
 Db 1113 TCGATAGGTTCTGAGACATC 1094

RESULT 9  
 US-09-270-542-87/c  
 ; Sequence 87, Application US/09270542  
 ; Patent No. 6322976  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Altman, Timothy  
 ; APPLICANT: Scott, James  
 ; APPLICANT: Stanton, Lawrence  
 ; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
 ; FILE REFERENCE: 4198/78179  
 ; CURRENT FILING DATE: 1999-03-17  
 ; EARLIER FILING DATE: 1999-03-17  
 ; EARLIER FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 207  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 87  
 ; LENGTH: 1709  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (540)..(550)  
 ; OTHER INFORMATION: The N at positions 540, 546, and 550 can be any  
 ; OTHER INFORMATION: nucleotide because the author is unsure of the  
 ; OTHER INFORMATION: exact sequence at these positions.  
 US-09-270-542-87

Query Match 69.1%; Score 15.2; DB 4; Length 1709;  
 Best Local Similarity 85.0%; Pred. No. 57;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATC 21  
 |||| |||||  
 Db 1113 TCGATAGGTTCTGAGACATC 1094

RESULT 10  
 US-09-270-542-85/c  
 ; Sequence 85, Application US/09270542  
 ; Patent No. 6322976  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Altman, Timothy  
 ; APPLICANT: Scott, James  
 ; APPLICANT: Stanton, Lawrence  
 ; TITLE OF INVENTION: Compositions and Methods of Disease Diagnosis and  
 ; FILE REFERENCE: 4198/78179  
 ; CURRENT FILING DATE: 1999-03-17  
 ; EARLIER FILING DATE: 1999-03-17  
 ; EARLIER FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 207  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 85  
 ; LENGTH: 2436  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-09-270-542-85

Query Match 69.1%; Score 15.2; DB 4; Length 2436;  
 Best Local Similarity 85.0%; Pred. No. 61;  
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATC 21

Db 1146 TCGATAGGTTCTGAGACATC 1127  
 |||| |||||  
 |||| |||||

RESULT 11  
 US-09-724-864-12/c  
 ; Sequence 12, Application US/09724864  
 ; Patent No. 6380362  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Watson, James D  
 ; APPLICANT: Murison, James G.  
 ; TITLE OF INVENTION: Polynucleotides, polypeptides expressed  
 ; FILE REFERENCE: 11000.105001  
 ; CURRENT FILING DATE: 2000-11-28  
 ; CURRENT FILING DATE: 2000-11-28  
 ; PRIOR FILING DATE: 1999-12-23  
 ; NUMBER OF SEQ ID NOS: 72  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 802  
 ; TYPE: DNA  
 ; ORGANISM: Mouse  
 ; ORGANISM: Mouse  
 US-09-724-864-12

Query Match 67.3%; Score 14.8; DB 4; Length 802;  
 Best Local Similarity 88.9%; Pred. No. 82;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 AACGGTTCGAGTCATC 22  
 |||| |||||  
 Db 399 AACAGGTCGAGTCATC 382

RESULT 12  
 US-09-134-001C-1980/c  
 ; Sequence 1980, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT FILING DATE: 1998-08-13  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 1980  
 ; LENGTH: 1248  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-1980

Query Match 66.4%; Score 14.6; DB 4; Length 1248;  
 Best Local Similarity 81.0%; Pred. No. 1.1e+02;  
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGAGTCATC 21  
 || |||||  
 Db 1140 ATTAACGGTTCGAGTCATC 1120

RESULT 13  
 US-09-149-476-208  
 ; Sequence 208, Application US/09149476  
 ; Patent No. 6420526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: 186 Human Secreted proteins

FILE REFERENCE: pz002p1	
CURRENT APPLICATION NUMBER: US/09/149,476	
CURRENT FILING DATE: 1998-09-08	
EARLIER APPLICATION NUMBER: PCT/US98/004493	
EARLIER FILING DATE: 1998-03-06	
EARLIER APPLICATION NUMBER: 60/040,162	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/040,333	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/038,621	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/040,626	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/040,334	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/040,336	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/040,163	
EARLIER FILING DATE: 1997-03-07	
EARLIER APPLICATION NUMBER: 60/047,600	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,615	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,597	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,502	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,633	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,583	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,617	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,618	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,503	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,592	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,581	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,584	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,500	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,587	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,492	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,598	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,613	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,582	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,596	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,612	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,632	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/047,601	
EARLIER FILING DATE: 1997-05-23	
EARLIER APPLICATION NUMBER: 60/043,580	
EARLIER FILING DATE: 1997-04-11	
EARLIER APPLICATION NUMBER: 60/043,568	
EARLIER FILING DATE: 1997-04-11	
EARLIER APPLICATION NUMBER: 60/043,314	
EARLIER FILING DATE: 1997-04-11	
EARLIER APPLICATION NUMBER: 60/043,569	
EARLIER FILING DATE: 1997-04-11	
EARLIER APPLICATION NUMBER: 60/043,311	
EARLIER FILING DATE: 1997-04-11	





Search completed: January 10, 2003, 06:37:34  
Job time : 3.85235 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 10, 2003, 01:01:36 ; Search time 2.87949 Seconds  
(without alignments)  
3362.497 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgaacggtcttgatcatct 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 389086 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	78.2	846	9	US-09-938-842A-421
2	17.2	78.2	1889	10	US-09-775-879-20
3	17.2	78.2	2925	10	US-09-775-879-22
4	15.8	71.8	1293	10	US-09-974-300-30
5	15.6	70.9	292	10	US-09-294-093B-461
6	15.6	70.9	2313	9	US-09-938-842A-2640
7	15.6	70.9	2447	10	US-09-823-114-12
8	15.6	70.9	3109	10	US-09-925-301-544
9	15.6	70.9	3954	9	US-10-071-766-44
10	15.4	70.0	4006	10	US-09-925-300-580
11	15.2	69.1	1469	10	US-09-917-800A-499
12	15.2	69.1	2484	10	US-09-747-835A-5
13	15.2	69.1	3188	10	US-09-747-835A-3
14	15.2	69.1	21252	10	US-09-070-927A-94
15	15.2	69.1	23934	10	US-09-764-860-777
16	15.2	69.1	23934	10	US-09-764-877-2336
17	15.2	69.1	23934	10	US-09-764-877-2544
18	15.2	69.1	70768	9	US-10-135-322-13
19	15	68.2	510	10	US-09-764-877-146

20	15	68.2	991	10	US-09-764-877-2302	Sequence 2302, Ap
21	15	68.2	1868	10	US-09-764-877-2301	Sequence 2301, Ap
22	14.8	67.3	990	9	US-09-938-842A-155	Sequence 155, App
23	14.6	66.4	258	10	US-09-923-876-2317	Sequence 2317, Ap
24	14.6	66.4	422	10	US-09-924-035A-476	Sequence 476, App
25	14.6	66.4	610	9	US-09-736-457-1094	Sequence 1094, Ap
26	14.6	66.4	610	9	US-09-902-941-1094	Sequence 1094, Ap
27	14.6	66.4	610	9	US-09-849-626-1094	Sequence 1094, Ap
28	14.6	66.4	655	10	US-09-770-149-502	Sequence 502, App
29	14.6	66.4	687	10	US-09-917-800A-1219	Sequence 1219, Ap
30	14.6	66.4	693	9	US-09-938-842A-1871	Sequence 1871, Ap
31	14.6	66.4	816	9	US-09-815-242-9194	Sequence 9194, Ap
32	14.6	66.4	1077	9	US-09-938-842A-1129	Sequence 1129, Ap
33	14.6	66.4	1189	10	US-09-900-237-5	Sequence 5, Appli
34	14.6	66.4	1473	10	US-09-070-844-25	Sequence 25, Appl
35	14.6	66.4	1506	10	US-09-070-844-23	Sequence 23, Appl
36	14.6	66.4	1583	12	US-10-062-254-273	Sequence 273, App
37	14.6	66.4	1686	9	US-09-938-842A-564	Sequence 564, App
38	14.6	66.4	1969	10	US-09-070-844-7	Sequence 7, Appli
39	14.6	66.4	2000	9	US-09-938-842A-4819	Sequence 4819, Ap
40	14.6	66.4	2096	10	US-09-070-844-19	Sequence 19, Appl
41	14.6	66.4	2099	10	US-09-070-844-3	Sequence 3, Appli
42	14.6	66.4	2137	10	US-09-070-844-18	Sequence 18, Appl
43	14.6	66.4	2140	10	US-09-070-844-1	Sequence 1, Appli
44	14.6	66.4	2319	9	US-10-121-032-14	Sequence 14, Appl
45	14.6	66.4	2733	9	US-09-935-868-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1

US-09-938-842A-421  
; Sequence 421, Application US/09938842A

; Patent No. US20020160378A1

; GENERAL INFORMATION:

; APPLICANT: Harper, Jeff

; APPLICANT: Kreps, Joel

; APPLICANT: Wang, Xun

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAININ

; TITLE OF INVENTION: SAME AND METHODS OF USE

; FILE REFERENCE: SCRIPT300-3

; CURRENT APPLICATION NUMBER: US/09/938, 842A

; CURRENT FILING DATE: 2001-08-24

; PRIOR APPLICATION NUMBER: US 60/227, 866

; PRIOR FILING DATE: 2000-08-24

; PRIOR APPLICATION NUMBER: US 60/264, 647

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/300, 111

; PRIOR FILING DATE: 2001-06-22

; NUMBER OF SEQ ID NOS: 5379

; SEQ ID NO 421

; LENGTH: 846

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-938-842A-421

Query Match 78.2%; Score 17.2; DB 9; Length 846;

Best Local Similarity 86.4%; Pred. No. 8.1;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTCTCGATCATCT 22

Db 56 ACCGAATCGTCTCGATCATCT 77

RESULT 2

US-09-775-879-20

; Sequence 20, Application US/09775879

; Patent No. US2002006822A1

; GENERAL INFORMATION:

; APPLICANT: Choe, Sunghwa

```
; APPLICANT: Feldmann, Kenneth A
; TITLE OF INVENTION: Dwf7 MUTANTS
; FILE REFERENCE: 2225-0003
; CURRENT APPLICATION NUMBER: US/09/775,879
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/179,901
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 1889
; TYPE: DNA
; ORGANISM: Genomic dwf7 (Arabidopsis)
US-09-775-879-20

Query Match      78.2%; Score 17.2; DB 10; Length 1889;
Best Local Similarity 86.4%; Pred. No. 9.1;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCTGAGTCATCT 22
    | |||| ||||| ||||| |||||
Db 198 ACCGAATCGTTCTGAGTCATCT 219

RESULT 3
US-09-775-879-22
; Sequence 22, Application US/09775879
; Patent No. US20020068822A1
; GENERAL INFORMATION:
; APPLICANT: Choe, Sunghwa
; APPLICANT: Feldmann, Kenneth A
; TITLE OF INVENTION: Dwf7 MUTANTS
; FILE REFERENCE: 2225-0003
; CURRENT APPLICATION NUMBER: US/09/775,879
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/179,901
; PRIOR FILING DATE: 2000-02-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2925
; TYPE: DNA
; ORGANISM: Genomic HDF7
US-09-775-879-22

Query Match      78.2%; Score 17.2; DB 10; Length 2925;
Best Local Similarity 86.4%; Pred. No. 9.8;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCTGAGTCATCT 22
    | |||| ||||| ||||| |||||
Db 1564 ACCGAATCGTTCTGAGTCACCT 1585

RESULT 4
US-09-974-300-30
; Sequence 30, Application US/09974300
; Patent No. US2002014621A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30

; APPLICANT: Bacillus licheniformis
US-09-974-300-30

Query Match      71.8%; Score 15.8; DB 10; Length 1293;
Best Local Similarity 89.3%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCTGAGTCA 19
    | |||| ||||| ||||| |||||
Db 369 ATCGACGGTTCTGATCA 387

RESULT 5
US-09-294-093B-461/c
; Sequence 461, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 461
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700342780H1
; NAME/KEY: unsure
; LOCATION: 266-267
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-461

Query Match      70.9%; Score 15.6; DB 10; Length 292;
Best Local Similarity 81.8%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCTGAGTCATCT 22
    | |||| ||||| ||||| |||||
Db 254 ATCAACGGGTTTGAGTAATCT 233

RESULT 6
US-09-938-842A-2640
; Sequence 2640, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2640
```

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05382
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 544
; LENGTH: 3109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1011)
; OTHER INFORMATION: n equals a,t,g; or c
US-09-925-301-544

Query Match 70.9%; Score 15.6; DB 10; Length 3109;
Best Local Similarity 81.8%; Pred. No. 66;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
||||| |||||| |||||
DB 386 ATCGAAGCTTCTGAATCATTT 407

RESULT 9
US-10-071-766-44
; Sequence 44, Application US/10071766
; Publication No. US20020192678A1
; GENERAL INFORMATION:
; APPLICANT: Huel-Mei Chen
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE
; FILE REFERENCE: PA-0043 US
; CURRENT APPLICATION NUMBER: US/10/071,766
; CURRENT FILING DATE: 2002-02-07
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: PERL Program
; SEQ ID NO 44
; LENGTH: 3954
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020192678A1 251859.2
; NAME/KEY: unsure
; LOCATION: 3786, 3788, 3791
; OTHER INFORMATION: a, t, c, g, or other
US-10-071-766-44

Query Match 70.9%; Score 15.6; DB 9; Length 3954;
Best Local Similarity 81.8%; Pred. No. 69;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATCGAAGGTTCTGAGTCATCT 22
||||| |||||| |||||
DB 1157 ATCGAAGCTTCTGAATCATTT 1178

RESULT 10
US-09-925-300-580
; Sequence 580, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300

```

```
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 580
; LENGTH: 4006
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-580

Query Match      70.0%; Score 15.4; DB 10; Length 4006;
Best Local Similarity 94.1%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 ACGGTTCTGAGTCATCT 22
    |||||
Db 237 ACGGTTCTGAGTCACCT 253
    |||||

RESULT 11
US-09-917-800A-499/c
; Sequence 499, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 499
; LENGTH: 1469
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AF072411
US-09-917-800A-499

Query Match      69.1%; Score 15.2; DB 10; Length 1469;
Best Local Similarity 85.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCTGAGTCATC 21
    |||||
Db 1097 TCGATAGGTTCTGAGACATC 1078
    |||||

; Sequence 5, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-1
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2484
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-747-835A-5

Query Match      69.1%; Score 15.2; DB 10; Length 2484;
Best Local Similarity 85.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CGAACGGTTCTGAGTCATCT 22
    |||
Db 1875 CGATGGTTCTGGGTCATCT 1894
    |||

RESULT 13
US-09-747-835A-3
; Sequence 3, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunrui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Asundi, Vinod
; APPLICANT: Dmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-1
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
```

; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: US 09/598,042  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: US 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: US 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3  
; LENGTH: 3188  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2484)  
US-09-747-835A-3

Query Match 59.1%; Score 15.2; DB 10; Length 3188;  
Best Local Similarity 85.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CGAACGGTTCGAGTCATCT 22  
||| ||||| |||||  
Db 1875 CGATGGTCTGGGTCACT 1894

RESULT 14  
US-09-070-927A-94  
; Sequence 94, Application US/09070927A  
; Patent No. US20020120116A1  
; GENERAL INFORMATION:  
; APPLICANT: Charles A. Kunsch  
; Patrick J. Dillon  
; Steven Barash  
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 982  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/070,927A  
; FILING DATE: 04-May-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/046,655  
; FILING DATE: 1997-05-16  
; APPLICATION NUMBER: 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: 60/066,009  
; FILING DATE: 1997-11-14  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kenley K. Hoover  
; REGISTRATION NUMBER: 40,302  
; REFERENCE/DOCKET NUMBER: PB369  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21252 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 94:  
US-09-070-927A-94

Query Match 69.1%; Score 15.2; DB 10; Length 21252;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
; QY 2 TCGAACGGTTCGAGTCATC 21  
||| ||||| |||||  
; Db 10687 TCAACAGTCTCTGATTCATC 10706

RESULT 15  
US-09-764-860-777/c  
; Sequence 777, Application US/09764860  
; Patent No. US20020094953A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008  
; CURRENT APPLICATION NUMBER: US/09/764,860  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1198  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 777  
; LENGTH: 23934  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-860-777

Query Match 69.1%; Score 15.2; DB 10; Length 23934;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATCGAACGGTTCGAGTCAT 20  
||| ||||| |||||  
Db 13230 ATAAAGGGTCTGAGTCAT 13211

Search completed: January 10, 2003, 06:43:17  
Job time : 8.87949 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 9, 2003, 21:14:32 ; Search time 74.5156 Seconds  
(without alignments)  
4781.561 Million cell updates/sec

Title: US-09-581-500B-13

Perfect score: 22

Sequence: 1 atcgacgggtcttgatcatct 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estinv:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	244	9	AA889105
2	22	100.0	270	9	AA904435
3	22	100.0	636	14	BQ807824
4	22	100.0	930	14	BQ435307
5	19.4	88.2	408	12	BG452389
6	19.4	88.2	529	12	BE999204

c	7	19.4	88.2	632	12	BF639518
c	8	19.4	88.2	653	10	BE318984
c	9	19.4	88.2	668	12	BG449328
c	10	17.8	80.9	715	12	BG467650
c	11	17.8	80.9	773	17	BH590723
c	12	17.2	78.2	143	17	AQ957346
c	13	17.2	78.2	298	12	BF899987
c	14	17.2	78.2	349	12	BF899989
c	15	17.2	78.2	396	12	BF738194
c	16	17.2	78.2	424	10	AW374777
c	17	17.2	78.2	425	12	BE932180
c	18	17.2	78.2	433	12	BF738261
c	19	17.2	78.2	447	12	BG008704
c	20	17.2	78.2	450	12	BG008684
c	21	17.2	78.2	456	17	AQ914000
c	22	17.2	78.2	458	10	AW062933
c	23	17.2	78.2	459	10	AW003072
c	24	17.2	78.2	459	10	AW962224
c	25	17.2	78.2	460	10	AW266225
c	26	17.2	78.2	460	10	AW374720
c	27	17.2	78.2	475	10	AW062934
c	28	17.2	78.2	489	10	AW374725
c	29	17.2	78.2	499	12	BF883208
c	30	17.2	78.2	509	17	AQ799516
c	31	17.2	78.2	544	9	AI670991
c	32	17.2	78.2	550	9	AJ285111
c	33	17.2	78.2	558	17	BH387822
c	34	17.2	78.2	598	17	BH383813
c	35	17.2	78.2	618	10	AW375534
c	36	17.2	78.2	669	14	BF002543
c	37	17.2	78.2	698	10	AW994568
c	38	17.2	78.2	701	12	BF295216
c	39	17.2	78.2	731	17	AQ258221
c	40	17.2	78.2	744	10	AW053439
c	41	17.2	78.2	749	17	CNS01E8P
c	42	17.2	78.2	782	12	BF529413
c	43	17.2	78.2	858	17	CNS01L30
c	44	17.2	78.2	901	12	BF964582
c	45	17.2	78.2	927	17	CNS01NOK

#### ALIGNMENTS

RESULT 1  
AA889105  
LOCUS AA889105 244 bp mRNA linear EST 04-JAN-1999  
DEFINITION am38h04.sl Soares\_NFL\_T\_GBC\_Sl Homo sapiens cDNA clone  
IMAGE:1471063 3', mRNA sequence.  
ACCESSION AA889105  
VERSION AA889105.1 GI:3015984  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 244)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)  
JOURNAL Tumor Gene Index  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
This clone is available royalty-free through LBNL; contact the IMAGE Consortium ([info@image.lbnl.gov](mailto:info@image.lbnl.gov)) for further information.  
Insert Length: 482 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 192.  
Location/Qualifiers  
1..244  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1471063"

FEATURES  
source  
/db\_xref="taxon:9606"  
/clone="IMAGE:1471063"

```

/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCL_CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
```

BASE COUNT 74 a 61 c 63 g 46 t

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 244;

Best Local Similarity 100.0%; Pred. No. 3.6;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22  
|||||

Db 121 ATCGACGGTTCGAGTCATCT 142

RESULT 2

AA904435 270 bp mRNA linear EST 09-JUN-1998

LOCUS ok07q11.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone

DEFINITION IMAGE:1507172 3', mRNA sequence.

ACCESSION AA904435

VERSION AA904435.1 GI:3039558

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

COMMENT Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 437 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 232.

Location/Qualifiers

1..270

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1507172"

/clone\_lib="Soares\_NFL\_T\_GBC\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site.1: Not I; Site.2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCL\_CGAP\_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo. "

BASE COUNT 77 a 66 c 74 g 53 t

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 270;

Best Local Similarity 100.0%; Pred. No. 3.8;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22  
|||||

Db 121 ATCGACGGTTCGAGTCATCT 142

RESULT 3

B0807824/c 636 bp mRNA linear EST 31-JUL-2002

LOCUS NISC.kk10f04.y1 NCL\_CGAP\_Brn72 Macaca mulatta cDNA clone

DEFINITION IMAGE:5331199 5', mRNA sequence.

ACCESSION B0807824

VERSION B0807824.1 GI:22032033

KEYWORDS EST.

SOURCE rhesus monkey.

ORGANISM Macaca mulatta

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

TITLE Cercopithecinae; Macaca.

1 (bases 1 to 636)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

cDNA Library Preparation:

CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: LLAM1840 row: K column: 8

Seq primer: M13R1 reverse primer (ABI).

Location/Qualifiers

1..636

/organism="Macaca mulatta"

/db\_xref="taxon:9544"

/clone="IMAGE:5331199"

/clone\_lib="NCL\_CGAP\_Brn72"

/tissue\_type="hypothalamus"

/note="Organ: brain; Vector: pCMV-SPORT6.cdb; Site.1:

NotI; Site.2: EcoRV; Cloned unidirectionally. Primer:

Oligo 67. Average insert size 2.2 kb. Constructed by

Invitrogen. Note: this is a NCI-CGAP Library."

BASE COUNT 81 a 210 c 190 g 155 t

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 636;

Best Local Similarity 100.0%; Pred. No. 5.4;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGACGGTTCGAGTCATCT 22  
|||||

Db 431 ATCGACGGTTCGAGTCATCT 410

RESULT 4

B0435307 930 bp mRNA linear EST 24-MAY-2002

LOCUS AGENCOURT\_7926752 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6160773

DEFINITION 5', mRNA sequence.

ACCESSION B0435307

VERSION B0435307.1 GI:21174383

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



```

REFERENCE 1 (bases 1 to 930)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13511 row: p column: 22
High quality sequence stop: 607.
FEATURES
source
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6160773"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 268 a 210 c 263 g 186 t 3 others
ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 930;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCGAAGCGTTCGAGTCATCT 22
|||||
Db 36 ATCGAAGCGTTCGAGTCATCT 57

RESULT 5
BG452389/c
LOCUS BG452389 408 bp mRNA linear EST 16-MAR-2001
DEFINITION NF097F021LF1016 Developing leaf Medicago truncatula cDNA clone
ACCESSION BG452389
VERSION BG452389.1 GI:13371183
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 408)
AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula leaf library
COMMENT Unpublished (2000)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 408 Std Error: 0.00
Plate: 097 row: F column: 02
Seq primer: TCACACGGAACACCTATGAC.
Location/Qualifiers
1..408
/organism="Medicago truncatula"
/db_xref="taxon:3880"

```

```

/clone="NF097F021LF"
/clone_lib="Developing leaf"
/tissue_type="leaf"
/dev_stage="Pooled developmental"
/Note="Vector: Lambda Zap; Contains a mixture of very
young, developing, mature and senescing leaves."
BASE COUNT 98 a 114 c 94 g 99 t 3 others
ORIGIN

Query Match 88.2%; Score 19.4; DB 12; Length 408;
Best Local Similarity 95.2%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGAGTCATCT 22
|||||
Db 242 TCGAACGGTTCGAGTCATCT 222

RESULT 6
BG999204/c
LOCUS BG999204 529 bp mRNA linear EST 06-OCT-2000
DEFINITION EST430927 GVSN Medicago truncatula cDNA clone pgVSN-15B23, mRNA
SEQUENCE.
ACCESSION BG999204
VERSION BG999204.1 GI:10699480
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 529)
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town
C.D., Bowman, C.L., Craven, M.B., Cho, J. and Fraser, C.M.
TITLE ESTs from senescent nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M273182e TIGR sequence name:
MTKBA127K More information is available at:
http://chrysie.tamu.edu/medicago
Seq primer: SKnod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
1..529
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pgVSN-15B23"
/clone_lib="GVSN"
/tissue_type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old
plants harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/Note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from the
mixture of effective nodules of 40 day old plants
harvested 36 hours post shoot removal and nodules
collected from 2 month old plants at mid-pod stage. The
cDNA was directionally ligated into the Uni-ZAP XR vector
from StrataGene and packaged using Gigapack III Gold
packaging extracts. Plasmids containing cDNA inserts were
excised from the recombinant lambda-ZAP phage using
Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT 122 a 119 c 145 g 143 t
ORIGIN

```

```

Query Match      88.2%; Score 19.4; DB 12; Length 529;
Best Local Similarity 95.2%; Pred. No. 85;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGTGAGTCATCT 22
    ||||| ||||| ||||| |||||
Db 69 TCGAACGGTTCGTGAGTCATCT 49

RESULT 7
BF639518/c      632 bp  mRNA  linear  EST 19-DEC-2000
DEFINITION      NF012C12IN1087 Insect herbivory Medicago truncatula cDNA clone
LOCUS            NF012C12IN 5', mRNA sequence.
ACCESSION        BF639518
VERSION           BF639518.1 GI:11903676
KEYWORDS          EST.
SOURCE            barrel medic.
ORGANISM          Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
AUTHORS          Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
                  ,H.R., Inman,J.F., Weller,J.W. and May,G.D.
TITLE            Expressed Sequence Tags from the Samuel Roberts Noble Foundation
                  Medicago truncatula insect herbivory library
JOURNAL           Unpublished (2000)
COMMENT           Contact: Korth K
                  Dept. of Plant Pathology
                  University of Arkansas
                  217 Plant Science Building, Fayetteville, AR 72701, USA
                  Tel: 501 575 5191
                  Fax: 501 575 7601
                  Email: kkorth@comp.uark.edu
                  Insert Length: 632 Std Error: 0.00
                  Plate: 012 row: C column: 12
                  Seq primer: TCACACGAGGAAACAGCTATGAC.
FEATURES
source            Location/Qualifiers
1..632
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF012C12IN"
/tissue_type="local and systemic leaves"
/dev_stage="mature"
/notes="Vector: Lambda Zap; Library was produced from fully
expanded M. truncatula leaves of plants fed upon by
Spodoptera exigua (beet armyworm) for 24 hours. Systemic
(undamaged leaves from injured plants) and wounded leaves
were harvested and pooled."
BASE COUNT        152 a 163 c 158 g 159 t
ORIGIN
1..632
TTCGACGGTTCGTGAGTCATCT 22
||||| ||||| ||||| |||||
Db 223 TCGAACGGTTCGTGAGTCATCT 203

RESULT 8
BE318984/c      653 bp  mRNA  linear  EST 21-DEC-2000
LOCUS            BE318984
DEFINITION        NF043F02L1F1016 Developing leaf Medicago truncatula cDNA clone
ACCESSION        BE318984
VERSION           BE318984.2 GI:11959620
KEYWORDS          EST.
SOURCE            barrel medic.

```

```

ORGANISM          Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
AUTHORS          Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
                  Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE            Expressed Sequence Tags from the Samuel Roberts Noble Foundation
                  Medicago truncatula leaf library
JOURNAL           Unpublished (2000)
COMMENT           On Jul 14, 2000 this sequence version replaced gi:9192761.
                  Contact: May GD
                  Plant Biology Division
                  The Samuel Roberts Noble Foundation
                  2510 Sam Noble Parkway, Ardmore, OK 73402, USA
                  Tel: 580 221 7391
                  Fax: 580 221 7380
                  Email: gdmay@noble.org
                  Medicago Genome Initiative accession: MGI:S:27143
                  Insert Length: 677 Std Error: 0.00
                  Plate: 043 row: F column: 02
                  Seq primer: TCACACGAGGAAACAGCTATGAC.
FEATURES
source            Location/Qualifiers
1..653
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF043F02LF"
/tissue_type="Developing leaf"
/dev_stage="Pooled developmental"
/notes="Vector: Lambda Zap; Contains a mixture of very
young, developing, mature and senescing leaves."
BASE COUNT        158 a 171 c 155 g 169 t
ORIGIN
1..653
TTCGACGGTTCGTGAGTCATCT 22
||||| ||||| ||||| |||||
Db 259 TCGAACGGTTCGTGAGTCATCT 239

Query Match      88.2%; Score 19.4; DB 10; Length 653;
Best Local Similarity 95.2%; Pred. No. 92;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGAACGGTTCGTGAGTCATCT 22
    ||||| ||||| ||||| |||||
Db 259 TCGAACGGTTCGTGAGTCATCT 239

RESULT 9
BG449328/c      668 bp  mRNA  linear  EST 16-MAR-2001
LOCUS            BG449328
DEFINITION        NF043G05IN1F1038 Insect herbivory Medicago truncatula cDNA clone
ACCESSION        BG449328
VERSION           BG449328.1 GI:13368109
KEYWORDS          EST.
SOURCE            barrel medic.
ORGANISM          Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
AUTHORS          Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
                  ,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE            Expressed Sequence Tags from the Samuel Roberts Noble Foundation
                  Medicago truncatula insect herbivory library
JOURNAL           Unpublished (2000)
COMMENT           Contact: Korth K
                  Dept. of Plant Pathology
                  University of Arkansas
                  217 Plant Science Building, Fayetteville, AR 72701, USA
                  Tel: 501 575 5191
                  Fax: 501 575 7601
                  Email: kkorth@comp.uark.edu
                  Insert Length: 668 Std Error: 0.00

```



Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 143)  
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,  
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.  
TITLE Genomic survey sequencing of *landsberg erecta* ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms  
JOURNAL Unpublished (2000)  
COMMENT Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: at@tigr.org  
For additional information, see <http://www.tigr.org/tdb/at/at.html>  
Seq primer: TF  
Class: Shotgun.

FEATURES  
source Location/Qualifiers  
1..143  
/organism="Arabidopsis thaliana"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"  
/clone="LERAP27"  
/clone\_lib="LERA"  
/note="Organ: leaf; Vector: pHC81; Total genomic DNA was  
sheared to 0.9-1 Kbp before ligation."  
BASE COUNT 47 a 24 c 15 g 57 t  
ORIGIN

Query Match 78.2%; Score 17.2; DB 17; Length 143;  
Best Local Similarity 86.4%; Pred. No. 5.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATCGAACGGTCTGAGTCATCT 22  
||||| ||| ||||| |||  
Db 64 ATCGACAGTTATGAGTCACT 43

RESULT 13  
BF899987  
LOCUS BF899987 298 bp mRNA linear EST 18-JAN-2001  
DEFINITION CM3-MT0195-051200-525-f04 MT0195 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF899987  
VERSION BF899987.1 GI:12291446  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 298)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-MT0195-051200-525-f04&t3=2000-12-05&t4=1>

Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 129.  
FEATURES  
source Location/Qualifiers  
1..298  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MT0195"  
/dev\_stage="Adult"  
/note="Organ: marrow; Vector: puc18; Site:1; SmaI; Site:2;  
SmaI; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No.196  
716 - Ludwig Institute for Cancer Research) profiles  
into the puc 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."  
BASE COUNT 77 a 77 c 66 g 78 t  
ORIGIN

Query Match 78.2%; Score 17.2; DB 12; Length 298;  
Best Local Similarity 86.4%; Pred. No. 7.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ATCGAACGGTCTGAGTCATCT 22  
||||| ||| ||||| |||  
Db 156 ACCGACAGTTCTGAGTCATTT 177

RESULT 14  
BF899989/c  
LOCUS BF899989 349 bp mRNA linear EST 18-JAN-2001  
DEFINITION CM3-MT0195-051200-525-h10 MT0195 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF899989  
VERSION BF899989.1 GI:12291448  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 349)  
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-MT0195-051200-525-h10&t3=2000-12-05&t4=1>  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 347.  
FEATURES  
source Location/Qualifiers  
1..349  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MT0195"  
/dev\_stage="Adult"  
/note="Organ: marrow; Vector: puc18; Site:1; SmaI; Site:2;  
SmaI; A mini-library was made by cloning products derived

BASE COUNT	105 a	111 c	84 g	96 t
ORIGIN				
Query Match		78.2%	Score 17.2;	DB 12; Length 395;
Best Local Similarity		86.4%;	Pred. NO. 8.3e+02;	
Matches 19;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;